Photoactivated Processes in Condensed Phase studied by Molecular Dynamics Simulations

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte

Dissertation

von

Lars Schäfer

aus Braunschweig

1. Referent: Prof. Dr. Marcus Elstner
2. Referent:Prof. Dr. Helmut Grubmüller
eingereicht am: 11.04.2007
mündliche Prüfung (Disputation) am: 13.06.2007
Druckjahr 2007

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

M. Andresen, M. C. Wahl, A. C. Stiel, F. Gräter, L. Schäfer, S. Trowitzsch, G. Weber, C. Eggeling, H. Grubmüller, S. W. Hell, and S. Jakobs: Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc. Nat. Acad. Sci. USA*, 2005, 102, 13070-13074.

O. F. Lange, L. Schäfer, and H. Grubmüller: Flooding in GROMACS: Accelerated barrier crossings in molecular dynamics. *J. Comp. Chem.*, 2006, 27, 1693-1702.

L. Schäfer, G. Groenhof, A. R. Klingen, G. M. Ullmann, M. Boggio-Pasqua, M. A. Robb, and H. Grubmüller: Photoswitching of the Fluorescent Protein asFP595: Mechanism, Proton Pathways, and Absorption Spectra. *Angew. Chemie int. Ed.*, 2007, 46, 530-536.

L. Schäfer, E. M. Müller, H. E. Gaub, and H. Grubmüller: Elastic Properties of Photoswitchable Azobenzene Polymers from Molecular Dynamics Simulations. *Angew. Chemie int. Ed*, 2007, 46, 2232-2237.

G. Groenhof, L. Schäfer, M. Boggio-Pasqua, M. Götte, H. Grubmüller, and M. A. Robb: Ultra-fast Deactivation of an Excited Cytosine-Guanine Base Pair in DNA. *J. Am. Chem. Soc.*, 2007, 129, 6812-6819.

Contents

Chapter 1

Introduction

Sunlight is an ubiquitous energy source that has enabled the development of life on earth. In the course of evolution, nature has developed elaborate bioenergetic mechanisms to convert solar energy into biologically usable chemical energy. The most prominent photobiological mechanism is photosynthesis, a light-driven process by which plants convert carbon dioxide and water into carbohydrates and molecular oxygen. At the molecular level, photosynthesis is a highly complex process involving many steps and a number of biomolecules, such as the light-harvesting complex, the photosynthetic reaction center, and the oxygen-evolving complex [1]. During photosynthesis, a pH gradient across a biomembrane is generated by the photosynthetic proteins. Also other biomolecules use light to generate a proton concentration gradient across a membrane, such as the the membrane protein bacteriorhodopsin, which acts as a light-driven proton pump [2]. The energy stored in this gradient can be used by the biological nanomachine ATP synthase to synthesize adenosine triphosphate (ATP), the generic energy-unit of the cell. Apart from its use as an energy source, organisms use light to gather information about their environment, for example in the vision process in animals, phototaxis in archaea and bacteria, and phototropy in plants.

These examples shall illustrate how the biological machinery makes use of the solar energy to build up essential molecules and to run vital processes. However, sunlight also contains significant amounts of harmful high-energy photons, such as ultraviolet (UV) light (wavelength ¡ 400 nm). These UV photons can destroy biomolecules. Solar light thus constitutes one of the most potent environmental carcinogens [3, 4, 5].

Deoxyribonucleic acid (DNA, Figure 1.1) carries the genetic information of all cellular forms of life, but due to the absorption of the DNA bases in the UV region of the spectrum, DNA is potentially vulnerable to structural damage induced by light. To protect the genetic information, highly elaborate mechanisms have evolved to tolerate or even repair damaged DNA [4]. More important, however, is the remarkable photostability of DNA, i.e., despite the absorption of a photon there is usually no structural damage [5]. This stability reduces the need for the energetically costly repair and might explain why DNA became the carrier of genetic information throughout the biosphere as a result of selection pressure during a long period of molecular evolution. Photostability arises from remarkably rapid deactivation pathways, which are only now coming into the focus of experiments (mainly through femtosecond laser spectroscopy) and theory. However, thus far very little is known about the dynamics underlying the mechanisms of DNA photostability.

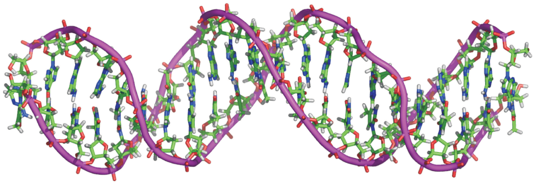


Figure 1.1: DNA double helix consisting of 22 base pairs.

Many organisms have evolved additional means to protect their genomes against the sun, such as the sea anemone *Anemonia sulcata* (Figure 1.2), which lives in shallow water and is thus frequently exposed to the sun. In its outer epithelial cells, the anemone expresses a protein named asFP595, a protein similar to the green fluorescent protein (GFP). The fluoroprotein asFP595 converts absorbed green light into a red fluorescence emission [6, 7, 8]. However, the fluorescence quantum yield is very low ($< 0.1\%$, Ref. [8]), and photoexcited asFP595 usually undergoes rapid radiationless deactivation without any structural damage [9]. Thus, the fluoroprotein functions as a highly efficient sun-blocker. By this means, the anemone further reduces the danger of DNA photodamage.

Figure 1.2: Photomontage of the sea anemone *Anemonia sulcata* with its fluorescent tentacles. The asFP595 fluoroprotein is modeled into the center of the anemone. The chromophore is shown as glowing spheres. The image of the anemone appears courtesy of Alexander Mustard.

Mankind has undertaken large efforts to follow nature's example and to make use of the energy stored in photons. Solar cells, for example, convert light into an electric current and in this respect resemble the biological proton pumps mentioned above. There have also been efforts to create artificial nanomachines that convert light (or other kinds of electromagnetic radiation) into mechanical work at the molecular level [10, 11, 12, 13]. The major reasons in favor of light are ease of addressability, picosecond reaction times to external stimuli, and compatibility with a broad range of ambient substances such as solvents, electrolytes, or gases. Therefore, nanomechanical devices or artificial molecular machines will, for a broad range of applications, most likely be powered by light, although the minimization of photodamage poses additional technical demands.

Understanding the molecular mechanisms underlying photochemical processes in complex systems such as a biomolecule or an artificial nanomachine is an intriguing and formidable task. Although remarkable progress has been made in the past decades, and despite the number and quality of available methods has tremendously increased, most mechanisms are poorly understood on a physical basis, which would require models based on first principles that allow a quantitative comparison with experimental results.

Experimental Techniques

Structural information is a prerequisite to study mechanisms at the molecular level. Experimental ensemble methods have made remarkable progress in determining the structure of biological macromolecules such as proteins or DNA (e.g., X-ray crystallography [14, 15] and nuclear magnetic resonance (NMR) [16, 17]) and, furthermore, even allow to probe dynamics (NMR relaxation [18, 19, 20], neutron scattering [21, 22], and electron paramagnetic resonance (EPR) [23]). At the single-molecule level, the atomic force microscopy (AFM)-based "mechanical triangulation" method has been recently developed, which allows to measure intramolecular distances with a sub-nanometer precision [24].

The concept that dynamics and function of most biomolecules are intimately coupled is now widely accepted [1, 25, 26, 27], but hard to prove. The molecular dynamics of many photoactivated processes are very fast, with typical timescales of nanoseconds to picoseconds or even faster. To track such ultra-fast atomic motion is a major challenge for experimental techniques. With typical time-resolutions of minutes (X-ray) or milliseconds (NMR), the conventional methods are too slow to capture fast photoactivated processes. More recently, advanced experimental techniques have been developed to follow fast processes. Single-molecule fluorescence spectroscopy techniques, such as single-pair fluorescence resonance energy transfer (spFRET) and single-molecule fluorescence polarization anisotropy (smFPA) can achieve time resolutions down to microseconds [28]. By measuring fluorescence intensity fluctuations, fluorescence correlation spectroscopy (FCS) can be used, e.g., to study diffusion processes or chemical reactions at the microseconds timescale [29]. Time-resolved X-ray diffraction allows to directly follow the conformational motion of biomolecules at the picoseconds timescale [30, 31], and time-resolved Fourier transform infrared spectroscopy (trFTIR) provides pico- to nanoseconds time resolution [32, 33, 34]. The trFTIR technique has for example been applied to study protein folding kinetics [35, 36, 37] as well as photochemical reactions in bacteriorhodopsin [38, 39] and photoactive yellow protein [40]. Time-resolved laser spectroscopy experiments have yielded very detailed information about biomolecules and their solvation dynamics even at the femtoseconds timescale [41, 42, 43, 44]. However, the wide-spread use of these techniques is impeded by the massive experimental effort involved. Furthermore, apart from X-ray techniques, the direct structural interpretation of these experiments is often difficult. Here, atomistic simulations can make significant contributions to the detailed understanding of molecular mechanisms.

Steady-state absorption and emission spectroscopy are classical experimental techniques to study photoactivated processes. However, these techniques generally do not provide any structural insight at the atomic scale. Femtochemistry and improved molecular beam techniques in conjunction with high-resolution spectroscopy [45, 46, 47, 48, 49, 50, 51, 52] have led to a renaissance of photochemical experiments. The successful measurement of the fluorescence lifetimes of various nucleosides is a representative example [51].

Theoretical Approaches

Compared to this tremendous experimental progress and the enormous variety of available techniques, theoretical approaches to study biological processes are still in their children's shoes. In particular, photoactivated processes in condensed phase have hitherto not been extensively studied due to computational and methodological bottlenecks. Molecular dynamics (MD) simulations yield the dynamics of molecular systems at a femtoseconds time resolution and, simultaneously, at an atomic spatial resolution. This level of resolution is well beyond what is currently achievable experimentally, and thus MD is a very promising method to study fast photoactivated processes in detail. As described in chapter 2, classical MD trajectories are generated by integrating Newton's equations of motion. The large number of pair-wise interactions to be calculated and the short time steps enforced by the fastest motions (such as bond- and angle-vibrations) lead to a very large computational effort. This currently limits atomistic MD to system sizes of about one million atoms and to processes that occur at timescales up to several hundred nanoseconds. Therefore, from a computational point of view, fast processes are easier to study than slow processes thanks to the smaller number of integration time steps.

However, the application of MD to photoexcited processes is not straightforward. As detailed in section 2.1, conventional MD simulations rely on empirical molecular mechanical (MM) force fields, which approximate the potential energy surface by a number of simple energy functions. On the one hand, this approximation enables to efficiently calculate the interactions between the atoms and thus drastically speeds up computation times, thereby rendering processes of up to several 100 ns accessible. On the other hand, by construction, most force fields cannot describe processes involving electronic reorganizations, such as electronic excitations, charge-fluctuations or -transfer, and the forming and breaking of chemical bonds. To accurately describe these processes requires quantum mechanical (QM) methods. However, the enormous size of a macromolecular system in water prohibits a complete quantum mechanical treatment during an MD simulation due to computational bottlenecks. To overcome this limitation, the hybrid QM/MM technique combines the advantages of both methods, i.e., the capability of QM to accurately treat electronic reorganizations with the computational efficiency of MM [53]. In the QM/MM approach, quantum mechanics is only used to describe the small subsystem which undergoes electronic reorganization, whereas the surrounding — usually by far the largest part of the overall system — is treated by molecular mechanics. For example, in a photochemical process, only the light-absorbing and emitting chromophore moiety is treated by a QM method, whereas a force field is used to model the surrounding. The QM/MM approach has been successfully applied to study, e.g., chemical reactions in enzymes and in solution [54, 55, 56, 57], surface reactivity [58, 59], zeolites [60, 61, 62], and, recently, the activation pathway of a photoreceptor protein [63].

In contrast to adiabatic reactions in the electronic ground state, photochemistry involves electronically excited states. To correctly model the dynamics of a photoactivated process requires the accurate description of both the ground and excited state energy surfaces. The enormous computational effort to obtain a reliable representation of electronically excited states has thus far obviated an extensive treatment of photochemical reactions in condensed phase by means of MD simulations, except from a few cases [63, 64, 65]. In addition, because photochemical reactions start in the excited state and end up in the ground state, it is essential to model the transitions ("hops") between the energy surfaces. Section 2.3.3 details how the concept of the conical intersection [66] can be used to achieve this goal.

Although being much more efficient than complete quantum mechanical MD simulations, QM/MM MD simulations are computationally significantly more demanding than conventional force field based simulations. This aggravates the problem of limited timescales (see above). A number of techniques have been devised to tackle this problem [67, 68]. However, most of them require as an input a reaction pathway, the knowledge of the energy hypersurface, putative intermediate states, or, at least, knowledge of the product. Given the complexity of biological processes, reaction pathways or intermediates usually cannot be guessed from chemical intuition, but rather need to be predicted from the simulation. Furthermore, due to the many degrees of freedom involved, the minimum potential energy pathway is never followed at 300 K, and, accordingly, free energies become relevant. A method addressing this problem is “flooding” [69, 70, 71], which is described in section 2.6.

Aims of the Present Thesis

The aim of the present thesis is to contribute to the understanding of photochemical processes in condensed phase by elucidating the molecular mechanisms of three different processes by means of MD simulations, the photoswitching mechanism of the fluorescent protein asFP595 (chapters 4, 5), the deactivation of a photoexcited DNA cytosine-guanine base pair (chapter 6), and the optical contraction of a photoswitchable polyazobenzene peptide (chapter 7). Spanning different fields from biology on the one end to nanotechnology on the other, each of these studies addresses particular questions, which are detailed at the beginning of the corresponding chapters, along with additional background information.

Mechanism of a Fluorescent Protein

The fluorescent protein asFP595 causes the fluorescence of the tentacle tips of the sea anemone *Anemonia sulcata* (Figure 1.2). Fluorescent proteins have been widely used as genetically encodable fusion tags to monitor protein localization and dynamics in living cells [72, 73, 74]. Interestingly, and in contrast to most other fluorescent proteins known this far, asFP595 fluorescence can be reversibly switched *on* and *off* in response to irradiation of a particular wavelength [6, 7, 8]. Thus, together with a few other proteins that have similar properties, these fluoroproteins have been termed optical highlighters. As the reversible photoswitching of photochromic organic molecules, such as fulgides or diarylethenes, is usually not accompanied by fluorescence [75], switching reversibility is a very remarkable and unique feature that may allow fundamentally new applications. Hence, not surprisingly, optical highlighters hold great promise in many areas of science beyond their prominent use as triggerable markers in live cells. For example, the reversible photoswitching, also known as kindling, may provide nanoscale resolution in far-field fluorescence optical microscopy much below the diffraction limit [76, 77, 78]. The group of Stefan Hell has recently demonstrated in a proof-of-principle fluorescence microscopy experiment that the photoswitching of asFP595 can yield spatial resolutions of about 50–100 nm in the focal plane, even at low laser intensities [79]. Since fluorescence can be sensitively read out from a bulky crystal, the prospect of erasable three-dimensional data storage is equally intriguing [80]. Currently, however, with the low quantum yield and rather slow switching kinetics, the photochromic properties of asFP595 leave much room for improvement. To systematically exploit the potential of such switchable proteins and to enable rational improvements to the properties of asFP595, a detailed molecular understanding of the switching process is mandatory. The aim of chapters 4 and 5 of the present thesis is to obtain detailed insights into the photoswitching mechanism of asFP595 at the atomic level. We suggest a complete and new mechanism that is grounded in a number of known experimental results and that contains predictions that are rigorously testable by experimental means.

Photostability of DNA Base Pairs

DNA carries the genetic information of all cellular forms of life. Figure 1.1 shows DNA in its usual double helix form, in which the nucleoside bases of the single strands are stacked upon each other, forming strong hydrogen bonds with the bases in the complementary strand (Watson-Crick configuration) [81, 82]. The nucleic acid bases are the primary absorbing constituents of DNA and can be excited by light in the harmful UV region of the spectrum. These photoexcited states are at the beginning of the complex chain of events that could culminate in photodamage [3, 5]. However, when arranged in the Watson-Crick (WC) configuration, the base pairs are remarkably photostable. Recent experiments by Abo-Riziq and co-workers indicate that the excited state of the cytosine-guanine (C-G) base pair has an extremely short lifetime in the order of a few tens of femtoseconds [83]. Potentially harmful photochemical processes such as the photodimerization of nucleobases typically occur at much slower timescales and thus cannot compete with such an ultra-fast deactivation [4]. Static ab initio calculations by Domcke and co-workers in vacuo suggest that ultra-fast deactivation may be triggered by a barrierless single proton transfer in the excited state [84, 85, 86]. As described in chapter 6, we have studied the origin of the intrinsic photostability of C-G base pairs *in vacuo* and also embedded in the DNA by means of excited state MD simulations.

Photoswitchable Polymers

Azobenzene is a well-studied photoactive system, which can be selectively photoswitched between an extended *trans* and a more compact *cis* conformation at two different wavelengths (Figure 1.3) [87, 88]. Incorporation of azobenzene units into a polypeptide backbone results in a polymer whose length can be optically lengthened and contracted, as demonstrated by recent single-molecule atomic force microscopy (AFM) experiments by Hermann Gaub and co-workers [89, 90].

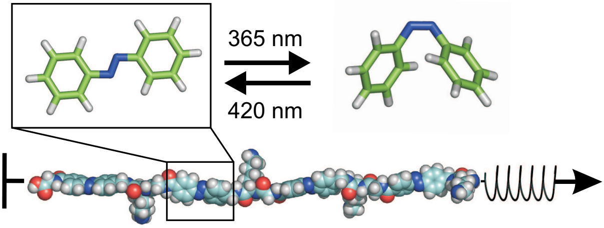


Figure 1.3: Azobenzene can be switched from *trans* to *cis* and back at 365 nm and at 420 nm, respectively (top). An optically contractible polymer can be generated by incorporating azobenzene moieties into the polymer backbone (bottom).

In analogy to the cycle of an Otto-engine, the switching could even be established in an optomechanical operating cycle, in which optical contraction against an external force delivered net mechanical work, thus demonstrating that photoswitchable polymers hold great promise for future applications in nanotechnology, e.g., as a light-triggered molecular switches or cargo lifters. However, work-output due to the overall optical contraction of the polymer was much smaller than expected. Chapter 7 reports the results of atomistic force-probe molecular dynamics simulations of photoswitchable azobenzene polymers. The simulation results are discussed in light of the AFM experiments and explain the observed elastic properties in terms of the underlying structural dynamics at the atomic scale.

Chapter 2

Theory and Concepts

This chapter outlines the general theoretical framework of the present thesis and the basics of the applied methods. It is subdivided into three parts. First, classical molecular dynamics simulations are introduced. Second, electronic structure methods that were used in this work are briefly described. Third, relevant concepts and their translation into algorithms are explained, such as conical intersections and diabatic surface hopping, QM/MM, force-probe MD, and flooding. The aim of this chapter is to selectively highlight those aspects that are of particular concern for the present work, and this chapter is by no means intended to be exhaustive. Further details on the particular methods and the simulation setup are given in the respective chapters. For a comprehensive description of molecular simulation techniques and electronic structure methods, see recent reviews [68, 91, 92] and text books [93, 94, 95, 96, 97].

2.1 Molecular Dynamics

Molecular dynamics (MD) simulations describe the evolution of a molecular system in time by numerically solving Newton's equations of motion. Conventional MD simulations can accurately describe the dynamics of large molecular systems of up to one million atoms and are based on three approximations: (i) nuclear and electronic motion are decoupled (Born-Oppenheimer approximation), (ii) nuclei behave as classical particles, and (iii) the interactions between the particles are described using an empirical force field. These three approximations are detailed below.

The Born-Oppenheimer Approximation

The evolution of a system in time is described by the time-dependent Schrödinger equation,

$$\widehat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}, \tag{2.1}$$

where $\widehat{H}$ is the Hamiltonian operator, i.e., the sum of the kinetic and potential energy operators, ψ is the wavefunction, and $\hbar$ is Planck's constant divided by 2π. The wavefunction ψ is a function of the coordinates and momenta of both, nuclei and electrons. The general idea behind the Born-Oppenheimer approximation [98] is to separate the fast degrees of freedom from the slow ones. In molecules, the nuclei and electrons can be considered the fast and slow degrees of freedom, respectively, because the nuclei are much heavier than the electrons. Hence, to a good approximation, nuclear and electronic motion can be decoupled, and the electrons can be considered to move in the field of fixed nuclei, i.e., the fast electrons instantaneously adjust to the slow nuclei. Consequently, the electronic wavefunction ψ_e depends only parametrically on the nuclear coordinates. The total wavefunction ψ_{tot} can thus be separated into an electronic and a nuclear part,

$$\psi_{\mathrm{tot}}\left(\mathbf{r},\mathbf{R}\right) = \psi_{\mathrm{n}}\left(\mathbf{R}\right)\psi_{\mathrm{e}}\left(\mathbf{r};\mathbf{R}\right), \tag{2.2}$$

where $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N)$ denotes the coordinates and momenta of the N nuclei and $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M)$ the coordinates and momenta of the M electrons, respectively. As the electrons move much faster than the nuclei, it is reasonable to approximate the electronic coordinates by their average values (averaged over the electronic wavefunction), and the nuclei move in the potential created by the average field of the electrons. This potential is called a potential energy surface (PES). Within the Born-Oppenheimer approximation, the nuclei move on a PES obtained by solving the (time-independent) electronic problem,

$$\widehat{H}_{\mathrm{e}}\psi_{\mathrm{e}} = E_{\mathrm{e}}\psi_{\mathrm{e}}, \tag{2.3}$$

where $\widehat{H}_{\mathrm{e}}$ is the electronic Hamiltonian and E_{e} is the energy eigenvalue, which parametrically depends on the nuclear positions $\mathbf{R}$. The Born-Oppenheimer approximation usually holds very well, except for a few special situations such as the crossing of two (or more) potential energy surfaces, as described below.

Classical Dynamics

For a typical macromolecular system with thousands of atoms, the solution of the time-dependent Schrödinger equation for the nuclear motion is prohibitively expensive. The central assumption in classical MD is that nuclei behave as classical particles, i.e., they obey Newton's equations of motion (Newton's second law)

$$-\nabla_i V\left(\mathbf{R}\right) = m_i \frac{d^2\mathbf{R}_i\left(t\right)}{dt^2}, \text{ or} \tag{2.4}$$

$$\mathbf{F}_i = m_i \mathbf{a}_i, \tag{2.5}$$

where $V\left(\mathbf{R}\right)$ is the potential energy, and $\mathbf{R}_i$ and m_i are the coordinates and mass of atom i, respectively. The force $\mathbf{F}_i$ acting on this atom determines its acceleration $\mathbf{a}_i$, which, within a discrete time step Δt, leads to a change of the velocity and position of the atom. The time step Δt has to be chosen small enough to capture the fastest motions in the system. These fastest motions are bond and angle vibrations, especially those involving hydrogen atoms, which occur at the femtosecond timescale and restrict Δt to about 1 fs. Integrating Newton's equations of motion is at the heart of an MD code. A number of algorithms have been devised to efficiently generate MD trajectories. In this work, the leap-frog modification of the Verlet scheme was applied [99],

$$\mathbf{R}\left(t + \Delta t\right) = \mathbf{R}\left(t\right) + \mathbf{v}\left(t + \frac{1}{2}\Delta t\right)\Delta t \tag{2.6}$$

$$\mathbf{v}\left(t + \frac{1}{2}\Delta t\right) = \mathbf{v}\left(t - \frac{1}{2}\Delta t\right) + \mathbf{a}\left(t\right)\Delta t, \tag{2.7}$$

where the current position $\mathbf{R}\left(t\right)$ and accelerations $\mathbf{a}\left(t\right)$ are stored together with the mid-step velocities $\mathbf{v}\left(t - \Delta t/2\right)$. This stable algorithm has the advantage that the expensive force calculation has to be done only once per integration time step. Under normal conditions, Newton's second law is a good approximation for macromolecular systems. However, quantum effects such as the behavior at low temperatures or the tunneling of hydrogen atoms can not be described.

Force Fields

Although numerous quantum chemical methods are available [95, 96] and very efficient yet highly reliable electronic structure methods have been developed [100, 101, 102, 103, 104, 105, 106, 107], the quantum chemical evaluation of the potential $V(\mathbf{R})$ by solving the electronic Schrödinger equation (eq 2.3) is currently still too expensive, rendering extended MD simulations of biomolecules in water unfeasible. Therefore, a third approximation is applied, and the total potential energy is expressed as a sum of simple and easy-to-compute analytical functions, which, in combination with a corresponding set of empirical parameters, make up the so-called molecular mechanical (MM) force field, e.g.,

$$V = \sum_{\text{bonds i}} V_B^i + \sum_{\text{bond angles j}} V_\gamma^j + \sum_{\text{impropers k}} V_{\text{imp}}^k + \tag{2.8}$$

$$\sum_{\text{dihedrals l}} V_D^l + \sum_{\text{pairs } \alpha,\beta} \left(V_{\text{Coul}}^{\alpha,\beta} + V_{\text{LJ}}^{\alpha,\beta} \right). \tag{2.9}$$

The individual energy terms are harmonic (V_B, V_γ, V_{imp}) or cosine (V_D) functions for the bonded interactions, or, for the non-bonded interactions, are motivated by physical laws. The van der Waals attraction and Pauli repulsion V_{LJ} is typically cast in the form of the Lennard-Jones (6,12) term,

$$V_{\text{LJ}}(\mathbf{R}; C_{12}, C_6) = \sum_{\text{pairs } \alpha,\beta} \left[\frac{C_{12}(\alpha,\beta)}{R_{\alpha,\beta}^{12}} - \frac{C_6(\alpha,\beta)}{R_{\alpha,\beta}^{6}} \right], \tag{2.10}$$

where the parameters C_{12} and C_6 are the repulsion and attraction coefficients, respectively. The electrostatic interaction between (partial) atomic charges is described by the Coulombic law

$$V_{\text{Coul}}(\mathbf{R}, q) = \sum_{\text{pairs } \alpha,\beta} \frac{q_\alpha q_\beta}{4\pi\epsilon_0\epsilon_1 \mathbf{R}_{\alpha,\beta}}, \tag{2.11}$$

where q_α is the partial charge of atom α. In these non-bonded functions (eqs 2.10 and 2.11), the interactions are calculated pair-wise. This avoids the tedious description of many-body effects and leads to a quadratic scaling of the computational cost with the number of atoms. The number of energy terms in eq 2.9, their exact functional forms, and individual parameters vary between different force fields. In this thesis, the OPLS all-atom [108] (asFP595, azobenzene polymers) and AMBER99 [109] (DNA) force fields were used. Other popular force fields for biomolecular simulations are, e.g., CHARMM [110] and GROMOS [111, 112]. Since all these force fields cover a variety of molecules (including, e.g., polypeptides, polysaccharides, nucleic acids, and lipids), each of them contains a large set of parameters. The parameter sets are usually determined by multi-dimensional fitting to experimental data and the results from quantum chemical calculations. Biomolecular systems are generally not in the gas phase, and data for the condensed phase (experimental and theoretical) are used whenever possible.

Since the parameters are determined by multi-dimensional fitting, a single parameter, e.g., the partial charge of a single atom, has no physical meaning but makes sense only in light of the force field as a whole. This in turn means that, strictly speaking, changing the values of single parameters or adding new parameters requires the reparametrization of the complete force field or at least of parts of it. However, such a procedure is not only vastly time-consuming but also does not guarantee a significantly better force field, because those properties crucial for the problem at hand will be strongly related to certain force field parameters and only weakly to others. As described in chapters 4 and 7 of the present thesis, we have used quantum chemical calculations to obtain the crucial parameters for the asFP595 and azobenzene chromophores, respectively, which were not included in the original OPLS force field.

2.2 Electronic Structure Methods

Solving the (time-dependent) Schrödinger equation exactly is impossible for complex many-electron systems. The aim of electronic structure methods is to yield approximate solutions for the electronic Schrödinger equation (eq 2.3). Knowing the (approximate) wavefunction and thus the energy of a molecule as a function of the nuclear coordinates enables to determine critical points on the energy surface (minima, transition states, higher order saddle points) and their energies, and, by calculating the forces "on-the-fly" during a molecular dynamics simulation, even allows to capture the evolution of a molecular system in time.

In this section, the multiconfigurational *ab initio* methods, semi-empirical methods, and density functional methods that were applied in the present work will be briefly described. These methods differ in the way and extent to which they capture electron correlation, that is, the correlated movement of electrons in a many-body system. As a start, the Hartree-Fock method is introduced, which provides the basis for the multiconfigurational and semi-empirical methods that are described in the subsequent paragraphs. In this work, the multiconfigurational CASSCF (section 2.2.4) and semi-empirical INDO/S methods (section

2.2.2) were applied to describe the electronically excited states of the asFP595 chromophore (chapters 4, 5) and of the cytosine-guanine base pair (chapter 6). In addition, time-dependent density functional theory (TDDFT) was applied to the asFP595 chromophore (section 2.2.3). To account for the polarization of the wavefunction due to the surrounding, the quantum chemical methods were applied within a QM/MM framework, as described in section 2.4.

2.2.1 Hartree-Fock Theory

The essence of the Hartree-Fock (HF) approximation is to replace the complicated many-electron problem by a one-electron problem in which the electron-electron repulsion is treated in an average way. Per definition, the HF energy does not contain electron correlation, because the correlated movement of electrons is not treated explicitly. The HF approach is a single-determinantal approach in which the many-electron wavefunction is expressed as one determinant, the so-called Slater determinant [96]. For an M-electron system, the determinant reads

$$\Psi\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\right) = \left(M!\right)^{-1/2} \begin{vmatrix} \chi_i\left(\mathbf{x}_1\right) & \chi_j\left(\mathbf{x}_1\right) & \cdots & \chi_k\left(\mathbf{x}_1\right) \\ \chi_i\left(\mathbf{x}_2\right) & \chi_j\left(\mathbf{x}_2\right) & \cdots & \chi_k\left(\mathbf{x}_2\right) \\ \vdots & \vdots & \cdots & \vdots \\ \chi_i\left(\mathbf{x}_M\right) & \chi_j\left(\mathbf{x}_M\right) & \cdots & \chi_k\left(\mathbf{x}_M\right) \end{vmatrix} . \tag{2.12}$$

The Slater determinant is constructed from occupying M one-electron spin orbitals $(\chi_i, \chi_j, \ldots, \chi_k)$, without specifying which electron is in which orbital. It is convenient to introduce a shorthand notation by only showing the diagonal elements,

$$\Psi\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\right) = \left|\chi_i\left(\mathbf{x}_1\right) \chi_j\left(\mathbf{x}_2\right) \cdots \chi_k\left(\mathbf{x}_M\right)\right\rangle . \tag{2.13}$$

The spin molecular orbitals (MOs) only depend on the coordinates of one electron,

$$\mathbf{x} = \begin{pmatrix} \mathbf{r} \\ \omega \end{pmatrix} , \tag{2.14}$$

which is composed of spatial $\mathbf{r}$ and spin ω variables. A spin MO $\chi_i\left(\mathbf{x}_i\right)$ is the product of a spatial MO $\psi_i\left(\mathbf{r}\right)$ and one of the spin functions $\alpha\left(\omega\right)$ or $\beta\left(\omega\right)$ (spin up or spin down, respectively). Slater determinants meet the requirement of the anti-symmetry principle (exclusion principle of Wolfgang Pauli), and incorporate exchange correlation, i.e., the correlated motion of electrons with parallel spins (Fermi hole). In contrast, the motion of electrons with opposite spins remains uncorrelated (no Coulomb hole), and single-determinantal wavefunctions are thus customarily referred to as uncorrelated wavefunctions [96].

The simplest wavefunction describing the ground state of an M-electron system is the single Slater determinant $|\Psi_0\rangle = |\chi_1\chi_2\cdots\chi_M\rangle$. According to the variational principle, the best wavefunction of this functional form is the one with the lowest energy eigenvalue $E_0 = \langle\Psi_0|\,\widehat{H}\,|\Psi_0\rangle$. Minimizing E_0 by varying the spin orbitals leads to the Hartree-Fock equation

$$\widehat{f}\,(i)\,|\chi\,(\mathbf{x}_i)\rangle = \epsilon_i\,|\chi\,(\mathbf{x}_i)\rangle \tag{2.15}$$

with the Fock operator

$$\widehat{f}\,(i) = \widehat{h}\,(i) + v^{HF}\,(i)\,. \tag{2.16}$$

The Fock operator is the sum of a core operator $\widehat{h}\,(i)$ that describes the kinetic and potential energy for attraction to the nuclei of a single electron,

$$\widehat{h}\,(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N}\frac{Z_A}{r_{iA}}, \tag{2.17}$$

and the Hartree-Fock potential

$$v^{HF}\,(i) = \sum_b \widehat{J}_b\,(i) - \widehat{K}_b\,(i)\,, \tag{2.18}$$

which is the average potential (mean field) experienced by electron i due to the other electrons, and the sum runs over all spin-MOs b. eq 2.18 expresses the electron-electron interaction in terms of the two-electron operators $\widehat{J}$ and $\widehat{K}$. The coulomb operator

$$\widehat{J}_b\,(i)\,|\chi_a\,(i)\rangle = \left[\int d\mathbf{x}_j\,|\chi_b\,(j)|^2\,r_{ij}^{-1}\right]|\chi_a\,(i)\rangle \tag{2.19}$$

represents the average local potential at $\mathbf{x}_i$ due to an electron in χ_b, and the exchange operator

$$\widehat{K}_b\,(i)\,|\chi_a\,(i)\rangle = \left[\int d\mathbf{x}_j\chi_b^*\,(j)\,r_{ij}^{-1}\chi_a\,(j)\right]|\chi_b\,(i)\rangle \tag{2.20}$$

involves exchange of two electrons. Because $\widehat{J}$ and $\widehat{K}$ depend on the spin orbitals, the Fock operator $\widehat{f}\,(i)$ depends on its eigenfunctions and eq 2.15 has to be solved iteratively using a so-called self-consistent field (SCF) method.

Within the restricted ansatz, the electron configuration of closed-shell systems is described by occupying each MO with two electrons, one with spin α and the other with spin β. The spatial parts of the spin orbitals are identical for each pair of electrons. Spin can be eliminated using the orthonormality of the spin functions ($\langle\alpha|\,\alpha\rangle = \langle\beta|\,\beta\rangle = 1$ and $\langle\alpha|\,\beta\rangle = \langle\beta|\,\alpha\rangle = 0$), which leads to HF equations that only depend on spatial coordinates,

$$\widehat{f}\,(i)\,|\psi\,(\mathbf{r}_i)\rangle = \epsilon_i\,|\psi\,(\mathbf{r}_i)\rangle\,. \tag{2.21}$$

To solve eq 2.21, the unknown molecular orbitals are expanded into a set of spatial basis functions $\{\varphi_\mu\,(\mathbf{r})\,|\mu = 1, 2, \ldots, K\}$ in a linear expansion called linear combination of atomic orbitals (LCAO),

$$\psi_i\,(\mathbf{r}_i) = \sum_{\mu}^{K} C_{iv}\varphi_\mu\,(\mathbf{r})\,. \tag{2.22}$$

Usual choices for the basis functions $\varphi_\mu\,(\mathbf{r})$ are Slater $(\exp\left[-\zeta\,|\mathbf{r}-\mathbf{R}|\right])$ or Gaussian $\left(\exp\left[-\alpha\,|\mathbf{r}-\mathbf{R}|^2\right]\right)$ functions. Because the atomic orbitals are fixed, the problem of calculating the HF molecular orbitals is replaced by varying the coefficients C_{iv} such that the minimal energy is obtained. This approach was suggested by Roothaan [113] and converts eq 2.21 into a set of algebraic equations that can be solved by standard matrix techniques:

$$\mathbf{FC} = \mathbf{SC}\epsilon. \tag{2.23}$$

The Roothaan eq 2.23 contains the matrix of expansion coefficients $\mathbf{C}$, the molecular orbital energies ϵ, the overlap matrix $\langle\psi_\mu|\,\psi_v\rangle$, and the Fock matrix $F_{\mu v} = \int d\mathbf{r}_1\varphi_\mu^*\,(1)\,\widehat{f}\,(1)\,\varphi_v\,(1)$, which is the matrix representation of the Fock operator (eq 2.16). Again, the Fock matrix can be separated into a one-electron (core) and a two-electron part,

$$F_{\mu v} = H_{\mu v}^{core} + G_{\mu v} \tag{2.24}$$

$$G_{\mu v} = \sum_{\lambda\sigma} P_{\lambda\sigma}\left[(\mu v|\lambda\sigma) - \frac{1}{2}\,(\mu\lambda|v\sigma)\right], \tag{2.25}$$

with the density matrix $P_{\lambda\sigma} = 2\sum_j^{M/2} C_{\lambda j}C_{\sigma j}$. The two-electron integrals in eq 2.25 are expressed in shorthand notation,

$$(\mu v|\lambda\sigma) = \int d\mathbf{r}_1 d\mathbf{r}_2\varphi_\mu^*\,(1)\,\varphi_v\,(1)\,r_{12}^{-1}\varphi_\lambda^*\,(2)\,\varphi_\sigma\,(2)\,. \tag{2.26}$$

Since the number of these two-electron integrals is large (formally N^4 where N is the number of basis functions), their evaluation dominates the computer time needed for a HF calculation.

2.2.2 Semi-empirical Methods

The goal of semi-empirical methods is to make Hartree-Fock theory less computationally expensive without necessarily sacrificing accuracy. Because the most time-consuming step in a HF calculation is the assembly of the two-electron integrals (eq 2.25), semi-empirical methods estimate the value of these integrals in an *a priori* fashion using a number of parameters. These parameters are obtained from experimental data or from high-level calculations. Thereby, electron correlation is implicitly included. Subsequently, the complete neglect of differential overlap method (CNDO) [114, 115] and the intermediate neglect of differential overlap method (INDO) [116] will be briefly introduced. Other popular semi-empirical methods are AM1 [102], PM3 [106, 107], and OM2 [103]. The differences between the numerous semi-empirical methods arise from different integral approximations.

CNDO

The CNDO method adopts the following approximations:

1. A minimal basis of one Slater-type function per valence orbital is applied. Only atoms having s and p functions are parameterized.
2. The overlap matrix is defined by $S_{\mu v} = \delta_{\mu v}$, where δ is the Krönecker delta.
3. For the two-electron integrals, $(\mu v|\lambda\sigma) = \delta_{\mu v}\delta_{\lambda\sigma}$ is applied, meaning that the only non-zero two-electron integrals are those that have the μ and v basis functions as identical orbitals on the same atom, and the λ and σ functions are also identical orbitals on the same atom, but the second atom can differ from the first.
4. The remaining two-electron integrals are approximated as $(\mu\mu|\lambda\lambda) = \gamma_{AB}$, where A and B are the atoms at which the basis functions μ and λ are centered, respectively. The parameter γ can either be computed explicitly from s-type functions (since γ depends only on the atoms A and B, $(s_A s_A|s_B s_B) = (p_A p_A|s_B s_B) = (p_A p_A|p_B p_B)$, etc), or calculated

from ionization potentials (IP) and electron affinities (EA) using the Pariser-Parr approximation [117, 118].

5. One-electron integrals for diagonal matrix elements are given as

$$\langle\mu|\,\widehat{h}\,|\mu\rangle = -\mathrm{IP}_\mu - \sum_{A=1}^{N} \left(Z_A - \delta_{Z_A Z_B}\right)\gamma_{AB}, \qquad (2.27)$$

   where μ is centered on atom B.

6. One-electron integrals for off-diagonal matrix elements are defined as

$$\langle\mu|\,\widehat{h}\,|v\rangle = \frac{1}{2}\left(\beta_A + \beta_B\right) S_{\mu v}, \qquad (2.28)$$

   with μ and v centered on atoms A and B, respectively. In the original CNDO parameterization, the β values were adjusted using experimental data.

The CNDO formalism drastically reduces the number of non-zero two-electron integrals compared to HF from (formally) N^4 to N^2. In addition, the N^2 integrals do not have to be evaluated by explicit integration, but can be obtained easily according to point 3, which vastly reduces the computational cost. However, as might be expected, the simplifications also bring along some drawbacks. For example, CNDO usually cannot predict accurate molecular structures and fails to distinguish between different orbitals on the same center and between different orbitals (or orbital orientations) on two centers (point 4).

INDO and INDO/S

The aim of chapter 4 of the present thesis is to predict the optical absorption spectra for a series of protonation states of the chromophore of the asFP595 fluoroprotein. To this end, the semi-empirical INDO/S method was applied, which is briefly described here. As a modification of the CNDO formalism, the INDO method provides a more accurate description of the electron-electron interactions on the same center (point 4, above). The key change with respect to CNDO is to use different values for the one-center two-electron integrals $(ss|ss)$, $(ss|pp)$, $(pp|pp)$, $(pp|p'p')$, and $(sp|sp)$. The integral values can be estimated from spectroscopic data, rendering the INDO method useful for modeling UV/Vis spectra.

Zerner and coworkers reparametrized INDO specifically for spectroscopic problems and termed their model ZINDO or INDO/S [119, 120]. The INDO/S parametrization has proven to accurately describe the separations between various electronic states for different compounds, ranging from organic molecules [119] to transition metal complexes [121] and metalloenzymes [122], as long as no Rydberg excitations are involved [123].

2.2.3 Density Functional Theory

In chapter 4, to complement the results obtained at the semi-empirical level and as an additional check, the optical absorption spectra of asFP595 were also calculated at the time-dependent density functional level (TDDFT, see below).

Density functional theory (DFT) is based on two fundamental theorems of Hohenberg and Kohn [124]. The first HK theorem states that the total ground state energy of an electron gas, which contains exchange and correlation contributions, is completely determined by the electron density $E_0 = E\left[\rho_0\left(\mathbf{r}\right)\right]$. In other words, ρ_0 determines the external potential, i.e., the number of electrons, the ground state wavefunction, and thus all other electronic properties. The minimum of the functional $E\left[\rho\left(\mathbf{r}\right)\right]$ is the ground state energy of the system. The second HK theorem shows that a variational principle exists to determine the density. However, the exact functional mapping the electron density to the energy is unknown, and it is the holy grail of DFT to find that functional. In practice, different DFT methods use different functional forms to approximate $\rho\left(\mathbf{r}\right)$.

Kohn and Sham [125] have shown furthermore how the many-body problem can be replaced by an equivalent set of self-consistent one-electron equations. The Kohn-Sham functional of the total electronic energy for a number of doubly occupied orbitals ψ_i can be written as

$$E^{KS}\left[\rho\left(\mathbf{r}\right)\right] = 2\sum_i \left\langle\psi_i\right| - \frac{\nabla^2}{2}\left|\psi_i\right\rangle + \int d\mathbf{r} V_{ne}\left(\mathbf{r}\right)\rho\left(\mathbf{r}\right) \tag{2.29}$$

$$+\frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}'\frac{\rho\left(\mathbf{r}\right)\rho\left(\mathbf{r}'\right)}{\left|\mathbf{r}-\mathbf{r}'\right|} + E_{XC}\left[\rho\left(\mathbf{r}\right)\right], \tag{2.30}$$

where V_{ne} is the external potential, i.e., the static contribution of the interaction energy between electrons and nuclei, $\rho\left(\mathbf{r}\right) = 2\sum_i \left|\psi_i\left(\mathbf{r}\right)\right|^2$ is the electron density, and $E_{XC}\left[\rho\left(\mathbf{r}\right)\right]$ is the exchange-correlation functional. In contrast to the other terms in eq 2.30, E_{XC} cannot be derived analytically, but rather has to be given *ad hoc*. The different functional forms and parameters for the exchange-correlation kernel account for the various DFT methods available today (see, e.g., ref [97] for an overview).

The ground state electron density can be obtained from eq 2.30 through the variational principle. In analogy to Hartree-Fock theory (cf. section 2.2.1), it eventually yields the Kohn-Sham equations

$$\widehat{h}^{KS} \left|\psi_i\right\rangle = \epsilon_i \left|\psi_i\right\rangle \tag{2.31}$$

with eigenvalues ϵ_i and effective one-electron Kohn-Sham operator

$$\widehat{h}^{KS} = -\frac{\nabla^2}{2} + V_{ne} + \int d\mathbf{r}' \frac{\rho\left(\mathbf{r}'\right)}{\left|\mathbf{r}-\mathbf{r}'\right|} + V_{XC} \tag{2.32}$$

that is similar to the Fock operator in wave mechanics (eq 2.16). In principle, if the exact exchange-correlation functional was known, DFT would be exact in that it fully accounts for the correlated movement of the electrons. The main advantage of DFT is that electron correlation is included at a low computational cost comparable to a Hartree-Fock calculation. Due to the introduction of a Slater determinant and molecular orbitals from a LCAO, eqs 2.31 and 2.32 are similar to the respective eqs 2.15 and 2.16 in HF theory.

TDDFT

The Hohenberg-Kohn theorems hold for stationary (time-independent) densities $\rho\left(\mathbf{r}\right)$. Runge and Gross formulated the analogous theorems for time-dependent densities $\rho\left(\mathbf{r}, t\right)$ [126]. Their first theorem states that $\rho\left(\mathbf{r}, t\right)$ uniquely determines the time-dependent external potential, and the second theorem ensures the existence of a procedure to determine $\rho\left(\mathbf{r}, t\right)$. Based on these theorems, the time-dependent Kohn-Sham equation

$$i\frac{\partial\psi_i\left(\mathbf{r}, t\right)}{\partial t} = \widehat{h}^{KS}\left(\mathbf{r}, t\right)\psi_i\left(\mathbf{r}, t\right) \tag{2.33}$$

was derived, which provides the basis for time-dependent DFT (TDDFT) methods.

In chapter 4 of the present thesis, linear response TDDFT was applied to calculate the excitations of the asFP595 chromophore. Linear response TDDFT methods [127, 128] employ the fact that the perturbation of the ground state electron density $\rho\left(\mathbf{r}, t\right)$ by a time-dependent electric field E with frequency $\omega\left(t\right)$ induces a variation of the dipole moment, $\delta\mu = \alpha E$. The frequency-dependent mean polarizability $\alpha(\omega)$ has poles at the excitation energies of the unperturbed system [97],

$$\alpha(\omega) = \sum_I \frac{f_I}{\omega_I^2 - \omega^2}. \tag{2.34}$$

In this sum-over-states expression, ω_I are the excitation energies $E_I - E_0$, f_I the corresponding oscillator strengths, and the sum is over all excited states I of the system. TDDFT usually yields electronic excitation energies that are accurate to within a few tens of an eV. However, errors tend to be larger for certain difficult situations, such as Rydberg states (due to the wrong asymptotic behavior of most exchange-correlation functionals [97]), charge transfer excitations (due to the locality of the functionals [129]), or transitions with substantial double excitation character [130], as is also discussed in closer detail in chapter 4.

2.2.4 Configuration Interaction and CASSCF

All QM/MM excited state MD simulations presented in chapters 5 and 6 of the present thesis rely on a CASSCF description of the different electronic states involved in the respective photoactivated processes. As described below, CASSCF is a multiconfigurational method that relies on the Hartree-Fock wavefunction as a reference wavefunction.

The Hartree-Fock ground state $|\Psi_0\rangle$ (eq 2.13) represents, after finding the set expansion coefficients that minimizes the energy, the best single-determinantal approximation to the electronic ground state of the system. However, the ground state Slater determinant, composed of occupying M one-electron spin orbitals, is only one of many determinants that can be constructed from the set of $2K > M$ spin MOs. In fact, the number of different single determinants is as large as $\begin{pmatrix} 2K \\ M \end{pmatrix}$, and the HF ground state is just one of these. Taking the HF ground state as the reference, other determinants can be formed by exciting electrons from occupied to unoccupied orbitals, yielding singly (S), doubly (D), triply (T), ..., M-tuply excited configurations. These many-electron wavefunctions, called configurational state functions (CSFs), can then be used as a basis in which the exact many-electron wave function $|\Phi_o\rangle$ can be expanded. The full configuration interaction (CI) wavefunction includes all possible excitations and can be written in a symbolic form [96]

$$|\Phi_0\rangle = c_0 |\Psi_0\rangle + \sum_S c_S |S\rangle + \sum_D c_D |D\rangle + \sum_T c_T |T\rangle + ... \tag{2.35}$$

where $|S\rangle$ represents the CSFs involving single excitations, $|D\rangle$ double excitations, etc. This full CI approach is exact within the basis set limit. However, since the number of configurations rapidly increases, full CI is computationally viable only for the smallest systems. Therefore, the number of configurations has to be restricted, which is the basic idea of the multiconfigurational self-consistent field (MCSCF) method [131, 132]. The MCSCF wave function is a truncated CI expansion,

$$|\Psi_{MCSCF}\rangle = \sum_I c_I \, |\Psi_I\rangle \,, \tag{2.36}$$

in which both the CI expansion coefficients c_I and the orbital coefficients in $|\Psi_I\rangle$ (eq 2.22) are simultaneously optimized using the variational principle. In the complete active space self-consistent field (CASSCF) method, full CI is performed within a subset of the molecular orbitals (the active space). Exciting n electrons in a subset of m molecular orbitals is denoted CASSCF(n, m). CASSCF can yield reliable energy surfaces not only for the ground state, but also for excited states. To obtain, e.g., the first excited singlet state S_1, the optimization procedure to find the optimal set of coefficients has to be constrained not to the lowest energy (ground state), but to the second lowest.

Section 2.2.1 has shown that a single-determinantal ansatz neglects the correlation between electrons of parallel spin. By using many Slater determinants to describe the wavefunction, a good portion of electron correlation is recovered. In particular, the CASSCF method includes to a large extent the so-called static correlation, i.e., the electron correlation arising from the fact that different determinants contribute with similar weights to the overall wavefunction (near-degeneracy effects). In contrast, the recovery of dynamical correlation that arises from the correlated motion of the electrons is limited by the size of the active space. Perturbation methods to correct for this drawback have been proposed, such as CASMP2 [133] and CASPT2 [134]. For these methods, however, analytical derivatives are not available, rendering MD simulations prohibitively expensive. Furthermore, for the topology of most excited state energy surfaces, static correlation is considered to be more important than dynamical correlation. The CASSCF method has been successfully applied to describe a number of excited state processes, such as surface crossings [135, 136]. To accurately predict absorption spectra, however, dynamical correlation needs to be included [137].

2.3 Photochemistry

Photons can interact with matter. In particular, if the energy of a photon matches the energy gap between two electronic states in a molecule, the photon can be absorbed and promote an electronic excitation from the lower level to the higher energy level. Concomitant with this excitation process, the electronic wavefunction of the molecule changes. The changing charge distribution means that the excitation goes along with a transition dipole moment μ. The transition probability per time unit for going from the initial state i to the final state f, $\lambda_{i\to f}$, can then be described by Fermi's golden rule,

$$\lambda_{i\to f} = \frac{2\pi}{\hbar} \left|\langle\Psi_f|\,\mu\,|\Psi_i\rangle\right|^2 \rho_f, \tag{2.37}$$

where $\langle\Psi_f|\,\mu\,|\Psi_i\rangle$ is the matrix element of the transition, and ρ_f is the density of final states.

Due to its higher energy, the excited state is metastable and, at some point in time, will relax back to the ground state. As shown in Figure 2.1, there are in principle three different relaxation mechanisms: fluorescence, phosphorescence, and radiationless decay. Which of these processes dominates, depends on the system and the nature of the transition.

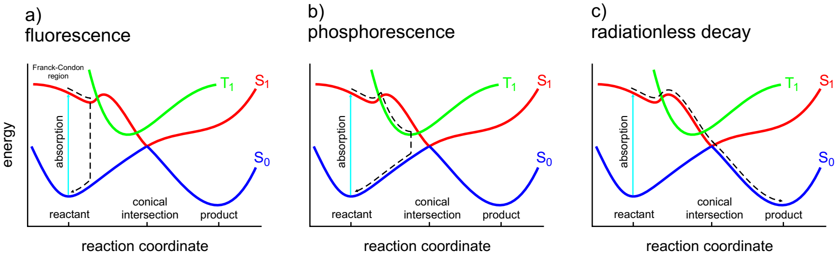


Figure 2.1: Possible excited state relaxation mechanisms. After photon absorption (cyan), the excited state S_1 (red) can relax back to the ground state S_0 (blue) via (a) fluorescence, (b) phosphorescence involving the triplet state T (green), or (c) decay at a conical intersection. In cases (a) and (b), a photon is emitted when the system returns back to S_0. In contrast, the decay is radiationless in case (c). Figure adopted from ref [138] with permission.

2.3.1 Fluorescence

The concept of fluorescence is illustrated in Figure 2.1a. Vertical excitation to the Franck-Condon region in the excited state S_1 is followed by rapid relaxation into a nearby local S_1 minimum. If the energy barrier in S_1 is too large to be overcome within nanoseconds, a fluorescence photon can be spontaneously emitted and the system returns to the ground state S_0. Due to the relaxation in S_1, the emitted photon is typically red-shifted with respect to the absorbed one. This effect is called the Stokes-shift.

2.3.2 Phosphorescence

For closed shell systems where all electrons are paired, the ground state is a singlet (S). The excited state wavefunctions, however, can be singlets or triplets (T), depending on whether all electrons are still paired or an electron spin has flipped upon excitation.[1] The conservation of overall momentum during photon absorption gives rise to certain selection rules for electronic transitions. The most important selection rule in this context is that S → T transitions are spin-forbidden. However, this does not strictly hold because of the coupling between the magnetic moment generated by the motion of the electrons with the spin magnetic moment. This process is called spin-orbit coupling. The presence of heavy nuclei such as sulfur can significantly enhance the probability of S → T transitions, since the fast motion of electrons around heavy nuclei leads to a stronger coupling.

In systems where singlet and triplet energy surfaces come energetically close or even intersect, the molecule might switch spin state by means of a process called intersystem crossing (IC), as shown in Figure 2.1b. Since the T_1 is lower than the S_1 minimum, the IC is irreversible, and the system is temporarily trapped in the triplet state. From the triplet minimum, the system can emit a photon and relax back to the ground state. Phosphorescence typically occurs at the microseconds to seconds timescale due to the spin-forbidden character of the transition. In this work, ultra-fast (sub-picosecond) excited state decay mechanisms (see next paragraph) were studied. Thus, ICs were neglected in our simulations.

---

[1] In principle, for double, triple, etc excitations, also higher multiplets such as quintets, septets, etc are possible. These transitions, however, usually correspond to higher energies and are thus not considered here.

2.3.3 Radiationless Decay at a Conical Intersection

There is a third possibility to relax to the ground state, which, in contrast to fluorescence and phosphorescence, is radiationless. Since this mechanism is particularly relevant for this work, it is explained in closer detail. Figure 2.1c shows that if the system encounters a point where the energy surfaces of the ground and the excited state cross, i.e., are energetically degenerate, radiationless decay back to the ground state can occur. Such a surface crossing is called a conical intersection (CI) [66, 136, 139, 140, 141, 142, 143] and is a highly efficient decay funnel back to the ground state. The topology of the CI and its location along a reaction coordinate determine whether either the initial ground state configuration is restored (photophysical process), or a chemically different species is formed (photochemical process, dashed line in Figure 2.1c). Photochemical reactions involve, for example, $2\pi + 2\pi$ cycloadditions, such as the photodimerization of DNA bases, photoisomerizations, as the 11-*cis* → all-*trans* retinal isomerization involved in the process of vision [64, 144, 145, 146] or the *trans-cis* isomerization of asFP595 (chapters 4 and 5), as well as changes in the hydrogen bonding pattern, like in the green fluorescent protein [147, 148, 149] or in DNA base pairs (chapter 6). The quantum yield of a photoactivated process is the ratio between the number of photons that lead to the formation of a desired photoproduct with respect to the number of absorbed photons.

In general, for a molecule with a set of electronic states $\{\Psi_k(\mathbf{r}, \mathbf{R})\}$, there is an energy gap between the states at an arbitrary nuclear conformation. In the Born-Huang approach [66], the total wavefunction for state k is expanded in a basis of M electronic wavefunctions,

$$\Psi_k(\mathbf{r}, \mathbf{R}) = \sum_{i=1}^{M} \Psi_i(\mathbf{r}; \mathbf{R})\, \psi_i^k(\mathbf{R}), \tag{2.38}$$

where $\psi_i^k(\mathbf{R})$ is the component of the overall nuclear wavefunction that corresponds to the ith electronic state $\Psi_i(\mathbf{r}; \mathbf{R})$. The electronic states are expanded into a large basis, e.g., according to eq 2.35. The Born-Huang approach is in principle correct, provided the set of the M CSFs is complete. However, as discussed in section 2.2.4, the expansion in eq 2.38 must be severely truncated.

To enable this truncation, the Ψ_k are chosen to be the adiabatic states, i.e., the eigenfunctions of the electronic Hamiltonian $\widehat{H}_\mathrm{e}(\mathbf{r}; \mathbf{R})$. Thus,

$$[H_e(\mathbf{r}, \mathbf{R}) - V_k(\mathbf{R})]\, \Psi_k(\mathbf{r}, \mathbf{R}) = 0, \tag{2.39}$$

where the eigenvalue $V_k(\mathbf{R})$ represents the kth potential energy surface. The secular equations

$$\left[\widehat{\mathbf{H}}(\mathbf{R})-\mathbf{V}(\mathbf{R})\right]\mathbf{c}(\mathbf{R})=0 \tag{2.40}$$

are obtained by substituting the equations 2.35 into the Schrödinger equation 2.39. The Hamiltonian in eq 2.40 is defined as

$$\widehat{H}_{k,l}(\mathbf{R})=\langle\Theta_k(\mathbf{r};\mathbf{R})|\,H_{\mathrm{e}}(\mathbf{r};\mathbf{R})\,|\Theta(\mathbf{r};\mathbf{R})\rangle_{\mathbf{r}}\,, \tag{2.41}$$

and $V_{k,l}(\mathbf{R})=V_k(\mathbf{R})\,\delta_{k,l}$. The CSFs Θ are constructed from the one-electron molecular orbitals, which can be determined from an MCSCF procedure, see section 2.2.4.

Consider a simple example in which eq 2.40 is formally evaluated in the two CSFs $\Theta_i(\mathbf{r})\,,\ i=1,2$. The Hamiltonian in eq 2.40 now becomes

$$\widehat{\mathbf{W}}(\mathbf{R})=\begin{pmatrix} H_{1,1}(\mathbf{R}) & H_{1,2}(\mathbf{R}) \\ H_{2,1}(\mathbf{R}) & H_{2,2}(\mathbf{R}) \end{pmatrix}=\mathbf{I}S(\mathbf{R})+\begin{pmatrix} -G(\mathbf{R}) & W(\mathbf{R}) \\ W(\mathbf{R}) & G(\mathbf{R}) \end{pmatrix}, \tag{2.42}$$

where $S(\mathbf{R})=(H_{1,1}+H_{2,2})/2$, $G(\mathbf{R})=(-H_{1,1}+H_{2,2})/2$, and $W(\mathbf{R})=H_{1,2}$. The Hamiltonian $\widehat{\mathbf{W}}$ has eigenvalues

$$V_\pm(\mathbf{R})=S(\mathbf{R})\pm\sqrt{\left(G(\mathbf{R})^2+WG(\mathbf{R})^2\right)} \tag{2.43}$$

and eigenfunctions

$$\begin{aligned}\Psi_1&=\Theta_1\cos\Lambda+\Theta_2\sin\Lambda\\ \Psi_2&=-\Theta_1\sin\Lambda+\Theta_2\cos\Lambda\end{aligned} \tag{2.44}$$

with

$$\tan 2\Lambda=\frac{W(\mathbf{R})}{G(\mathbf{R})}. \tag{2.45}$$

The eigenfunctions in eq 2.44 are the adiabatic wavefunctions at the CI. For the nuclear configuration $\mathbf{R}_{CI}$ to be a point of intersection, the energies of both states have to be equal. Thus, the degenerate solution for eq 2.43 exists only under the conditions $H_{1,1}=H_{2,2}$ and $H_{1,2}=0$, or

$$G(\mathbf{R}_{CI})=W(\mathbf{R}_{CI})=0. \tag{2.46}$$

To simultaneously fulfill these conditions requires that the two terms are independent, i.e., they are functions of different (uncoupled) nuclear coordinates. In diatomic molecules with only one internal degree of freedom, the two conditions can never be simultaneously satisfied for states of the same symmetry. This leads to the non-crossing rule of von Neumann and Wigner [150], which states that crossings can occur only if the states belong to different irreducible representations, implying $H_{1,2} = 0$. Most polyatomics are not symmetric, and all states thus belong to the same irreducible representation. However, the number of degrees of freedom is so large that the crossing conditions can always be fulfilled irrespective of the symmetry.

Two independent coordinates are needed to lift the degeneracy at a crossing. This can be shown by expanding the matrix elements around R_{CI}, the nuclear configuration at the intersection, in a Taylor series truncated at first order,

$$(H_{1,1} - H_{2,2}) = (H_{1,1} - H_{2,2})\,|_{\mathbf{R}_{CI}} + \nabla\,(H_{1,1} - H_{2,2})\,|_{\mathbf{R}_{CI}}\Delta\mathbf{R}$$

$$H_{1,2} = H_{1,2}|_{\mathbf{R}_{CI}} + \nabla H_{1,2}|_{\mathbf{R}_{CI}}\Delta\mathbf{R}, \tag{2.47}$$

where $\Delta\mathbf{R} = \mathbf{R} - \mathbf{R}_{CI}$. Truncation at first order is valid if only small steps in nuclear configuration space are taken. Since at $\mathbf{R}_{CI}$, $H_{1,1} - H_{2,2} = 0$ and $H_{1,2} = 0$, only the linear terms from equations 2.47 remain. Thus, in the vicinity of the CI, the there are two coordinates that lift the degeneracy, the gradient difference vector (GDV) $\mathbf{x}_1 = \nabla\,(H_{1,1} - H_{2,2})\,\Delta\mathbf{R}$, and the derivative coupling vector (DCV) $\mathbf{x}_2 = \nabla H_{1,2}\Delta\mathbf{R}$. As shown in Figure 2.2a, the surfaces in the branching space defined by the GDV and the DCV form a double-cone, because changing the nuclear configuration in the branching space linearly increases the energy gap. Along all other $N^{int} - 2$ internal degrees of freedom ($N^{int} = 3n - 6$ for an n-atom system), the degeneracy remains. Thus, the conical intersection is an $N^{int} - 2$-dimensional seam or hyperline in the configurational space spanned by the full number of degrees of freedom. As shown in chapter 6, for the special case of the cytosine-guanine base pair also $N^{int} - 1$-dimensional conical intersections exist. Typically, polyatomic molecules have a large number of atoms, leading to high-dimensional and thus ubiquitous surface crossing seams. Figure 2.2 illustrates the topology of a CI in subspaces spanned by different coordinates for the simple case of a hypothetical triatomic hydrogen-like molecule [139]. For $n = 3$, there are three internal degrees of freedom, the two bond lengths x_1 and x_2, and the bond angle α. Thus, the CI seam is one-dimensional, and the other two coordinates define the branching space.

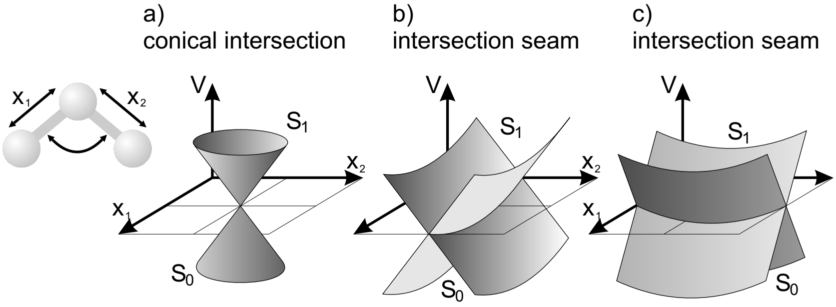


Figure 2.2: Conical intersection topologies in different configurational subspaces, illustrated for a hypothetical triatomic molecule (left). (a) In the branching space spanned by the gradient difference vector (GDV) and the derivative coupling vector (DCV), the two energy surfaces S_1 and S_0 form a double-cone with the CI as a point in the apex. (b,c) The subspace contains the orthogonal degree of freedom α, making the CI a one-dimensional line parallel to α. x_1 and x_2 are at their CI values in plots b and c, respectively. Figure adopted from ref [138] with permission.

For the hypothetical triatomic considered here, for the sake of simplicity, the branching space shall be spanned by the two bond lengths x_1 and x_2, which therefore represent the GDV and the DCV, respectively. In general, the branching space coordinates are more complex linear combinations of the internal degrees of freedom. Figure 2.2a shows the S_1 and S_0 energy surfaces forming a double-cone in the branching space, with the degeneracy in the apex. In Figure 2.2b, the surfaces are plotted in the subspace spanned by the bond angle α and x_2, while x_1 is fixed at the value corresponding the the CI geometry. Thus, the CI is now a one-dimensional line parallel to α. Figure 2.2c shows the two surfaces in the subspace spanned by x_1 and α, with x_2 being fixed at the conical intersection. In this case, only x_1 lifts the degeneracy, and the CI seam again lies parallel to α.

Figure 2.3 shows two important consequences of the adiabatic wavefunctions (eq 2.44). First, the two wavefunctions Ψ_1 and Ψ_2 at points on a circle close to the apex of the double-cone are diabatically related. As an example, we start on the upper S_1 surface at Ψ_2 close to the CI and increment $\Lambda \rightarrow \Lambda + \pi$. The S_1 wavefunction at $\Lambda + \pi$ is identical to the S_0 wavefunction at Λ, except for the sign:

$$\Psi_2\left(\Lambda+\pi\right)=-\Psi_1\left(\Lambda\right). \tag{2.48}$$

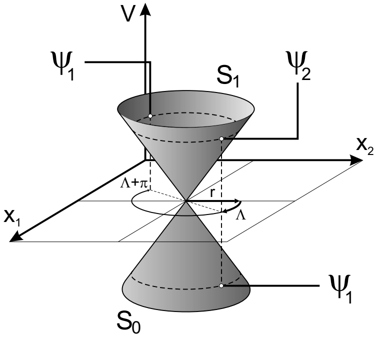


Figure 2.3: Adiabatic representation of a conical intersection in the branching space. The adiabatic wavefunctions Ψ at points on a circle around the CI are diabatically related (see text).

For eq 2.48 to hold, the radius r needs to be small such that the points on the circle are close to the CI. The diabatic relation of the wavefunctions through the conical intersection provides the key for a practical surface hopping algorithm for non-adiabatic MD simulations, as detailed below. The second consequence of the eigenfunctions is that moving around the CI in a full circle ($\Lambda \rightarrow \Lambda + 2\pi$) changes the sign of Ψ (Figure 2.3). Thus, obviously, the adiabatic wavefunction has a singularity at the CI, i.e., Ψ is not a single-valued function of the nuclear coordinates. This singularity results from the Born-Oppenheimer approximation and violates the basic postulate of quantum mechanics that a wavefunction has to be a single-valued function of its variables. The nuclear wavefunction $\psi\left(\mathbf{R}\right)$ ensures that the total wavefunction is indeed a proper single-valued function and thus rescues the problem.

Born-Oppenheimer Breakdown at the Conical Intersection

Close to a CI, the electronic states are nearly degenerate, and the energy gaps between different vibrational levels of the nuclear wavefunction are comparable in size to the gap between the electronic states. Thus, nuclear and electronic wavefunctions might resonantly interfere, and slight changes in the nuclear positions can induce transitions between the different electronic surfaces. In other words, different electronic states are coupled through the nuclear motion, a phenomenon which is called non-adiabatic coupling [151]. At a conical intersection, this non-adiabatic coupling is very strong, and two (or more) offsprings can emerge from a single nuclear wavepacket that initially belonged to a single electronic state [152]. In principle, the offsprings might coherently superpose. However, they evolve on different energy surfaces and, over time, diverge and acquire different phases. In high-dimensional polyatomic systems such as a biomolecule in condensed phase, the probability of coherence is negligible outside the non-adiabatic coupling regions, a process known as quantum decoherence [152]. In the non-adiabatic molecular dynamics simulations presented in the present work, the integration time step is typically in the order of a femtosecond, and a surface crossing is passed in a single MD step. Therefore, quantum coherence can be ignored.

Diabatic Surface Hopping

In our excited state molecular dynamics simulations, the nuclei are propagated classically on Born-Oppenheimer surfaces by calculating the forces on-the-fly from an CASSCF wavefunction (section 2.2.4) and integrating Newton's equations of motion (eq 2.5) using the leap frog algorithm (eqs 2.6 and 2.7). This approach is valid as long as the energy gap between different energy surfaces is large. If two states come close, the system can change from one state to the other. In the simulations described in this work, this hopping between different electronic states occurs at the conical intersection seam, as illustrated in Figure 2.4. A surface hop from S_1 to S_0 occurs if (i) the energy gap is below a defined threshold, and (ii) the dot-products of the CASSCF configuration interaction eigenvectors of the involved states indicate that a surface crossing seam has been encountered. The eigenvector of state I is composed of the expansion coefficients in eq 2.36. At MD step i, the system is on the excited state surface, and the vector c_2^i is optimized to calculate the gradient (Figure 2.4). Between steps i and $i+1$, the CI was passed. Thus, at step $i+1$ the excited state inner product $c_2^i c_2^{i+1}$ becomes small, whereas the mixed excited-ground state inner product $c_2^i c_1^{i+1}$ is close to one. If in addition the energy gap between the two states is below a given threshold, a surface hop took place. Now, the gradient is computed from S_0 instead of from S_1. Note that the gradients have to be calculated twice at step $i + 1$, because it can only be determined afterwards whether the system has to be propagated on S_1 or on S_0. This diabatic surface hopping approach is valid if the integration time step is small enough to ensure the diabatic relation between the wavefunctions before and after the CI. In our MD simulations, we typically use integration time steps of 0.5 or 1 fs, which are on the one hand small enough to ensure the diabatic relation of the wavefunctions, and on the other hand large enough to justify the neglect of quantum coherence (see above). Since the energy surfaces are degenerate at the crossing seam, diabatic surface hopping ensures energy conservation during the dynamics. This diabatic surface hopping algorithm was developed by Gerrit Groenhof and coworkers and implemented into the GROMACS/Gaussian QM/MM interface [63, 138].

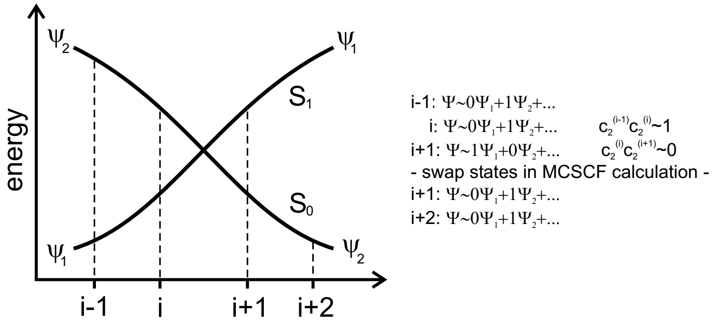


Figure 2.4: Diabatic surface hopping algorithm with surface selection at the conical intersection. At MD step i, the system is in the excited state, close enough to the CI that the states are diabatically related. The forces are calculated on-the-fly from the CASSCF wavefunction of the excited state S_1, which is dominated by Ψ_2 (right). The CI vector $c_2^{i-1} c_2^{i}$ is close to one. Between steps i and $i + 1$, the CI was passed. Since the wavefunctions are diabatically related around the CI, the excited state wavefunction is now dominated by Ψ_1, whereas Ψ_2 represents the ground state. Thus, the CI vector $c_2^{i} c_2^{i+1}$ is close to zero and $c_2^{i} c_1^{i+1}$ approaches one. A surface hop is assumed, and the gradient at step $i + 1$ is recomputed on S_0, the surface on which the simulation is subsequently continued. Figure adopted from ref [138] with permission.

2.4 QM/MM

The aim of this work is to describe the excited state dynamics of condensed phase systems. However, molecular mechanical (MM) force fields cannot accurately describe electronic reorganizations, such as electronic excitations, charge-fluctuations or -transfer, and the forming and breaking of chemical bonds. To treat such processes requires quantum mechanical (QM) methods, such as, e.g., the CASSCF method used in this work. However, QM calculations are computationally very demanding, rendering the complete QM treatment of large biological macromolecules prohibitively expensive. Luckily, for photoactivated processes in condensed phase, photon absorption is typically localized at a small part of the overall system, such as the chromophore of a fluoroprotein or a DNA base pair. Only this part requires a QM method, and an MM description suffices for the rest of the system. As already briefly introduced in chapter 1, such a partitioning is employed in a QM/MM scheme. In this subsection, the methodological basics of the QM/MM schemes applied in the course of the present work are briefly introduced.

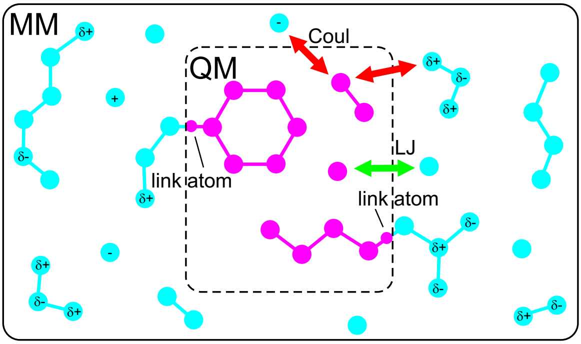


Figure 2.5: Basic idea of QM/MM. The overall system is subdivided into different regions. Only a small part of the overall system is described QM (magenta atoms), whereas the rest is modeled by an MM force field (cyan atoms). The open valences resulting from bonds across the QM/MM boundary are capped with link atoms. Non-bonded electrostatic (red arrows) and van der Waals interactions (green arrow) between the two regimes are included (see text).

The QM/MM approach was originally proposed by Warshel and Levitt [53] and is illustrated in Figure 2.5. The overall QM/MM Hamiltonian can be described as the sum of the individual terms plus a coupling term,

$$\widehat{H} = \widehat{H}_{QM} + \widehat{H}_{MM} + \widehat{H}_{QM/MM}. \tag{2.49}$$

The difficult part is the coupling (third term on the r.h.s. of eq 2.49), and the numerous QM/MM schemes in the literature arise from differential couplings between the QM and MM subsystems. In principle, there are three types of QM/MM schemes, which capture the non-bonded interactions between the two regions to different extents [153], (i) mechanical embedding, (ii) electronic embedding, and (iii) polarized embedding.

In the mechanical embedding (ME) scheme, the QM subsystem is embedded into the MM environment and sterically fixed by classical MM-type interactions. Two different approaches exist for the non-bonded interactions between the two regions. In the crudest approximation, the electrostatic interactions between the two regions are neglected, and the coupling is solely sterical *via* van der Waals interactions, which are described at the force field level (eq 2.10). Such mechanical embedding schemes, e.g., the ONIOM method by Morokuma and coworkers [154], have been successfully applied in cases where electrostatics are not important for the studied process [155] — a scenario that, however, does not apply for most biomolecules. A different suite of mechanical embedding schemes captures the electrostatic interactions between the two regions at the force field level. Atomic partial charges of the QM atoms are derived from the wavefunction through charge-fitting, and the electrostatic interactions are then described classically using the Coulombic law (eq 2.11).

The electronic embedding scheme explicitly includes the polarization of the wavefunction due to the charges of the surrounding MM atoms into $\widehat{H}_{QM}$. To this end, the charges are included into the one-electron operators, which now read (cf. eq 2.17)

$$\widehat{h}_{QM/MM}\,(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N} \frac{Z_A}{r_{iA}} - \sum_{MM\;atoms} \frac{q_\alpha}{r_{i\alpha}}, \tag{2.50}$$

where q_α is the (partial) charge of MM atom α, $r_{i\alpha}$ is the distance between electron i and MM atom α, and the second sum runs over all MM atoms. The interactions between the QM nuclei and the charged MM atoms are described by eq 2.11 and are added subsequently.

In the electronic embedding scheme, the wavefunction is polarized by the surrounding MM atoms. However, the reverse polarization (response) of the MM atoms to the changed wavefunction is neglected if common non-polarizable force fields are used. Thus, strictly speaking, electronic embedding is not internally consistent. In fact, the use of MM point charges and the neglect of back-polarization can in certain cases lead to an overpolarization of the wavefunction. This problem can be partly overcome by the use of gaussian-shaped charges instead of point charges [156], as was also done for the TDDFT QM/MM calculations of the absorption energies of asFP595 (chapter 4).

The polarized embedding scheme additionally allows polarization of the MM region in response to the changing wavefunction, thus correcting for the somewhat unbalanced coupling of the electronic embedding scheme. The original QM/MM scheme of Warshel and Levitt [53] already included such back-polarization. Some of the currently available QM/MM programs include polarized embedding implementations [157, 158, 159]. Since the mutual polarization of the two regions has to be calculated in a self-consistent manner, polarized embedding is significantly more costly than the other schemes. Thus, in this work, an electronic embedding scheme similar to the one originally proposed by Field and coworkers [160] was used.

An additional difficulty in treating the QM/MM coupling arises if the boundary between the two regions cuts through a bond, as is illustrated in Figure 2.5. In such a case, the following interactions are modeled at the force field level: the bond at the boundary, angles involving two MM atoms, and dihedrals involving at least two MM atoms. Furthermore, the open valence of the QM subsystem needs to be capped. The most straight forward approach, which was also followed in this work, is to introduce a hydrogen link atom at the QM/MM boundary (Figure 2.5) [161]. The link atom is invisible for the MM subsystem, and the forces acting on it are redistributed over the two atoms of the original bond according to the lever rule. A number of alternative approaches to cap the QM subsystem have been developed, such as the frozen orbital method [162, 163] or the generalized hybrid orbital method [164].

2.5 Force-Probe Molecular Dynamics

The major bottleneck of today's atomistic molecular dynamics simulations is that due to the enormous computational effort involved, only processes at nanoseconds to microseconds time scales or faster can be studied directly. This sampling problem is even aggravated in QM/MM simulations, where the costly quantum chemical calculation restricts the accessible timescales by at least three orders of magnitude as compared to conventional force field MD. Unfortunately, apart from a few exceptions, relevant processes, such as chemical reactions or many large scale conformational transitions in proteins or other (bio-)polymers, occur at slower time scales and therefore are currently far out of reach for MD. The force-probe molecular dynamics (FPMD) technique [165, 166] addresses this problem by probing the dynamics under an external force. This external force accelerates transitions and thus allows to probe energy barriers and reaction pathways. In this work, FPMD was applied to probe the elastic properties of a photoswitchable polymer, as described in chapter 7.

In FPMD, a harmonic spring is attached to one (or more) atoms i of the simulation system,

$$V_{spring,i}\left(t\right) = \frac{k_0}{2}\left[z_i\left(t\right) - z_{spring}\left(t\right)\right]^2, \tag{2.51}$$

where k_0 is the force constant of the spring, z_i the center of mass of the pulled atoms, and z_{spring} the position of the spring. The spring is then moved with constant velocity v along a predefined direction, $z_{spring}\left(t\right) = z_i\left(0\right) + vt$. Due to the moving spring, the pulled atoms experience an additional force

$$F_i = -k_0\left[z_i\left(t\right) - z_{spring}\left(t\right)\right]. \tag{2.52}$$

FPMD simulations closely mimic atomic force microscopy (AFM) or related experiments, such as optical or magnetic tweezers, and simulations have proven to be an invaluable tool for the interpretation of such experiments at the molecular level [167]. For example, FPMD has been successfully applied to study the stiffness of carbohydrates [168], protein-ligand unbinding [165], or the partial unfolding of proteins [169].

2.6 Flooding

The flooding technique addresses the sampling problem by inclusion of a flooding potential into the force field. This flooding potential locally destabilizes the educt state and thereby significantly accelerates the escape from the initial energy well without affecting the reaction pathway.

The implementation and testing of the flooding method in the official version 3.3 of the freely available MD program package GROMACS was done in close collaboration with Oliver Lange, who did the major part of the coding. I focused on two examples that shall demonstrate the application of flooding to accelerate conformational transitions and chemical reactions (chapter 3). The second example was carried out within a QM/MM framework. Furthermore, in chapter 5, flooding was used to accelerate the escape from an excited state minimum in a QM/MM simulation.

Examples for slow processes which involve a barrier crossing from an initial (educt) to a final (product) free energy well (Figure 2.6, top) are collective conformational transitions or activated chemical reactions, either in the ground or in the excited state. These rare events are often not directly accessible to conventional MD, impeding the elucidation of transition pathways. Many techniques to address this problem have been devised (for an overview, see refs [67, 170]), the majority of which rely on *a priori* knowledge of both, the educt *and* the product state. Therefore, if only the educt state is known, these methods cannot be applied, and the prediction of product states becomes a real challenge. This scenario is addressed by the flooding technique [69, 70], which aims at predicting both, the unknown product state as well as the transition pathway. The basic principle of flooding is described in Figure 2.6.

A number of related techniques to accelerate transitions by destabilizing potentials have been proposed, such as local elevation [171], hyperdynamics [172, 173] or boosted dynamics [174], and metadynamics [175]. These, however, have not been implemented in GROMACS and are therefore not considered here.

In the framework of the present work, I will briefly summarize the theory [69] and detail the implementation. Subsequently, user interface and file formats are described. Finally, the two examples presented in chapter 3 will serve to illustrate the application of the method and the interpretation of the results. The first example is the acceleration of the *trans-gauche* conformational transition of n-butane. The second example is a chemical reaction, the opening of a three-membered ring in aqueous solution. The latter example also demonstrates that flooding can be used within the QM/MM framework. Here, the benefit of using the flooding method is particularly pronounced, since chemical reactions are often connected with tremendous energy barriers. Furthermore, the mandatory application of vastly time consuming QM methods shortens the accessible time scales drastically. Here, particularly large acceleration can be gained using the flooding technique. Finally, most of the conventional methods mentioned above cannot be applied to reactions in condensed phase such as done in the second example.

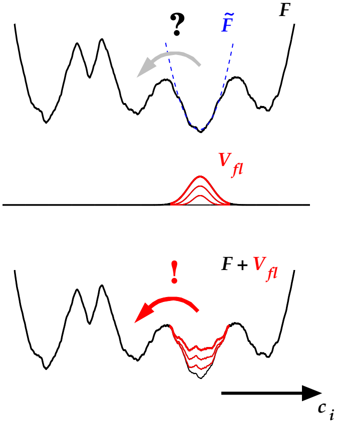


Figure 2.6: Principle of flooding. The free energy F along a collective coordinate, c, is approximated quasi-harmonically in the local educt minimum to yield $\tilde{F}$. From this, a gaussian-shaped flooding potential V_{fl} is constructed which destabilizes the initial well and accelerates the transition across the barrier.

2.6.1 Theory

First the conceptual framework is summarized, as detailed in references [69, 70], to clarify notation (Figure 2.6). Flooding involves two steps: First, the free energy landscape of the system is approximated quasi-harmonically [69]. Second, a multivariate flooding potential V_{fl} is constructed from this approximation, which serves to raise the bottom of the educt energy well without affecting regions of higher energy and, in particular, the barriers surrounding the energy well, which determine the transition pathway.

For the quasi-harmonic approximation of the free energy landscape, m linear collective coordinates

$$c_j = \sum_{i}^{3N} a_{ij} x_i \quad j = (1, \ldots, m)$$

are chosen. The x_i denote the $3N$ Cartesian coordinates of the molecule and the $3N \times m$ coefficients a_{ij} form an orthogonal matrix $\mathbf{A}$, which defines the collective coordinates $\mathbf{c} = (c_1, \dots, c_m)^{\mathrm{T}}$. Here, we consider two widely used methods to obtain such collective coordinates, principal component analysis (PCA) [176, 177, 178] and normal mode analysis (NMA) [179, 180, 181]. However, the theory is also valid for other suitably chosen coordinates.

The quasi-harmonic approximation in the collective coordinates and centered at $\mathbf{c}^{(0)} = \langle \mathbf{c} \rangle$,

$$\tilde{F}(c_1, \dots, c_m) = \frac{1}{2} k_{\mathrm{B}} T \sum_{j=1}^{m} \lambda_j \left(c_j - c_j^{(0)} \right)^2 ,$$

is chosen such that thermal motions within $\tilde{F}$ have the same amplitudes as the original atomic motion within the original well of the atomistic energy landscape. Accordingly, the curvatures λ_j are determined by the fluctuational amplitudes

$$\lambda_j^{-1} = \alpha^2 \left\langle \left(c_j - c_j^{(0)} \right)^2 \right\rangle ,$$

which can, e.g., be obtained from molecular dynamics. Here, the parameter α was included to allow global scaling of the quasi-harmonic free energy approximation, e.g., to account for insufficient sampling (see end of next chapter).

A gaussian-shaped flooding potential

$$V_{\mathrm{fl}}(c_1, \dots, c_m) = E_{\mathrm{fl}} \exp \left[-\frac{k_{\mathrm{B}} T}{2 E_{\mathrm{fl}}} \sum_{j=1}^{m} \lambda_j \left(c_j - c_j^{(0)} \right)^2 \right] , \qquad (2.53)$$

is constructed such that its principal axes are parallel to those of $\tilde{F}$, and its extensions (standard deviations) along these axes are proportional to those of the thermal fluctuations in $\tilde{F}$. The flooding strength E_{fl} controls both, the width as well as the height of the flooding potential.

If the harmonic approximation is sufficiently good, e.g., for small molecules in vacuo, or under certain conditions also for solvated proteins [182], the curvatures λ_j of $\tilde{F}$ can also be obtained from the vibrational frequencies ω_j. This fact can be used to replace MD sampling by NMA, which is computationally more efficient.

Two strategies can be used to determine the flooding strength. The first, rather straightforward strategy employs a *constant flooding* strength $E_{\text{fl}} \equiv E_{\text{fl}}^{(0)}$ throughout the simulation [69]. In complex systems, however, the relation between $E_{\text{fl}}^{(0)}$ and the expected acceleration of the transition is often unclear. To this end, the second strategy, *adaptive flooding* [70], couples the flooding strength E_{fl} to the flooding energy V_{fl} with the aim to reach and maintain a specified target destabilization free energy $\Delta F_0 = -k_{\text{B}}T \ln \left\langle e^{-V_{\text{fl}}/k_{\text{B}}T} \right\rangle \approx \langle V_{\text{fl}} \rangle$ [69], where the angle brackets denote an appropriate running time-average. From this free energy, the expected acceleration is obtained via the Boltzmann factor $e^{\Delta F_0/k_{\text{B}}T}$. The coupling is achieved by updating the flooding strength at each timestep i,

$$E_{\text{fl}}^{(i+1)} := E_{\text{fl}}^{(i)} + \frac{\Delta t}{\tau_1} \left[\Delta F_0 - \Delta F^{(i)} \right],$$

and taking a sliding average for $\langle V_{\text{fl}} \rangle$,

$$\Delta F^{(i+1)} := \left(1 - \frac{\Delta t}{\tau_2} \right) \Delta F^{(i)} + \frac{\Delta t}{\tau_2} V_{\text{fl}}^{(i)}.$$

Here, Δt is the integration step size used for the molecular dynamics simulation, and the time constants τ_1 and $\tau_2 = f\tau_1$ $(f \geq 1)$ define the coupling strengths.

2.6.2 Implementation

Overview

The flooding method was implemented into the program `mdrun`, which is the main molecular dynamics engine of the MD package GROMACS [183]. To take the additional forces due to V_{fl}, $F_i = -\nabla_i V_{\text{fl}}$, into account, the force routine of `mdrun` calls our flooding implementation within the `edsam` [184] module of GROMACS. All necessary information is condensed in a single input file, `sam.edi`, which is an extension of the old `edsam` input file. `sam.edi` is generated by means of the program `make_edi`. Since it is an ascii file, it can also be edited manually. To perform flooding simulations with GROMACS, `sam.edi` is passed to `mdrun` via the option `-ei`. The option `-eo` chooses a filename for the ascii output, e.g., `sam.edo`, which contains, in that order: the MD time step, E_{fl}, V_{fl}, and ΔF.

Input Generation

The flooding input file, `sam.edi`, which contains all necessary information (such as the flooding matrix $\mathbf{A}$ and parameters) is generated by the program `make_edi`. The required collective coordinates and the motional amplitudes can be obtained, e.g., by principal component analysis (PCA), normal mode analysis (NMA), or full correlation analysis (FCA).

To use PCA, the program `g_covar` computes the covariance matrix of fluctuational motion from an MD trajectory $\mathbf{x}(t)$. `g_covar` removes rotational and translational motion by least squares fitting to a reference structure $\mathbf{x}_{\mathrm{ref}}$, yielding a corrected trajectory $\mathbf{x}'(t)$. The covariance matrix $\mathbf{C} = \left\langle (\mathbf{x}' - \langle \mathbf{x}' \rangle)(\mathbf{x}' - \langle \mathbf{x}' \rangle)^{\mathrm{T}} \right\rangle$, where the brackets denote the time average, is diagonalized and its eigenvectors and eigenvalues are stored together with $\mathbf{x}_{\mathrm{ref}}$ and the average structure $\mathbf{x}_{\mathrm{av}} = \langle \mathbf{x}' \rangle$ in the files `eigenvec.trr` and `eigenval.xvg`. `make_edi` reads these files and stores m selected eigenvectors a_{ij} and the corresponding reciprocal eigenvalues σ_j, $\lambda_j = \sigma_j^{-1}$, as well as the reference and average structure in the flooding input file `sam.edi`.

NMA can be performed with `mdrun`, yielding `hessian.mtx`, which is subsequently diagonalized by `g_nmeig`. Note, that for the application of NMA within flooding no mass weighting should be applied [70]. The normal modes and their frequencies are also stored in files `eigenvec.trr` and `eigenval.xvg`. In order to treat the two cases NMA and PCA in a common framework, as done also further below, `make_edi` transforms the frequencies ω_j to curvatures according to $\lambda_j = \omega_j^2/k_{\mathrm{B}}T$. This also allows for one common format of the flooding input file.

In the following, a short overview of the command line interface of `make_edi` is given. The default behavior is to assume that the input files, provided with options `-f` for the vectors and `-eig` for the eigenvalues or frequencies, refer to PCA modes, created with `g_covar`. If the input files refer to normal modes, this has to be flagged by `-hesse`. The option `-flood <list>` selects m principal or normal modes. Note that the first six normal modes correspond to rotational and translational motion. These are (within numerical accuracy) zero and should not be used. Additional command line options are used to set parameters such as those which control flooding strength and adaptive flooding (cf. Table 2.1).

parameter	command line option
$E_{\text{fl}}^{(0)}$	-Eflnull
α	-alpha
τ	-tau
ΔF_0	-deltaF0

Table 2.1: Additional command line options for `make_edi`.

Coordinate Transformation

Within each flooding simulation time step, the forces due to the flooding potential V_{fl} need to be computed and added to the forces derived from the original atomistic potential. Computing the forces by directly evaluating $\frac{\partial V_{\text{fl}}}{\partial \mathbf{x}}$ is inefficient, however, mainly due to the $O\left(N^2\right)$ operations required for computing the matrix product $\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\Lambda\mathbf{A}\mathbf{x}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$. Instead, we apply the chain rule and compute

$$\mathbf{f}_{\text{fl}} = \frac{\partial V_{\text{fl}}}{\partial \mathbf{x}} = \frac{\partial V_{\text{fl}}}{\partial \mathbf{c}} \frac{\partial \mathbf{c}}{\partial \mathbf{x}}. \tag{2.54}$$

The benefit of this is that the gradient can be computed efficiently by evaluating the m equations

$$\frac{\partial V_{\text{fl}}}{\partial c_j} = \frac{k_{\mathrm{B}}T}{E_{\text{fl}}} \lambda_j V_{\text{fl}}(c_1, \ldots, c_m) \left(c_j - c_j^{(0)}\right)$$

plus a single evaluation of eq (2.53) to gain $V_{\text{fl}}(c_1, \ldots, c_m)$. Carried out in this way, the complete force evaluation scales with $O(Nm)$.

To project the atomic coordinates $\mathbf{x}$ to collective coordinates, first the rotational and translational motion are removed. Consequently, the center of mass $\mathbf{x}_{\text{cm}}$ is subtracted and a rotation $\mathbf{R}$ is applied, which is obtained by least squares fitting to the reference structure $\mathbf{x}_{\text{ref}}$: $\mathbf{x}' = \mathbf{R}\left(\mathbf{x} - \mathbf{x}_{\text{cm}}\right)$. The m collective coordinates are obtained by

$$\mathbf{c} = \mathbf{A}^{\mathrm{T}}\left(\mathbf{x}' - \mathbf{x}'_{\text{av}}\right),$$

where the average structure $\mathbf{x}'_{\text{av}}$ and the column vectors a_{ij} of the matrix $\mathbf{A}$ are read from the input file `sam.edi` during initialization of the MD program `mdrun`. Accordingly, the second part of the chain rule eq (2.54) reads

$$\frac{\partial \mathbf{c}}{\partial \mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{R}.$$

Restraining Potentials

The original idea of flooding is to destabilize conformations by applying the gaussian-shaped flooding potential V_{fl}. Additionally, a simple inversion of its sign allows to stabilize, i.e., restrain the system at an arbitrary position $\mathbf{c}^{(0)}$ of the collective coordinates. This sign inversion is switched by the option `-restrain` of the input file generator `make_edi`. The effect is accomplished by inverting the signs of the parameters α^2 and E_{fl} in the input file `sam.edi`. Note that the adaptive coupling to a target energy ΔF_0 can still be applied using inverted potentials, and might be judiciously used to gradually switch the restraints on or off. Furthermore, the switch `-harmonic` chooses a harmonic potential (rather than a Gaussian function)

$$V_{\mathrm{fl}} = -\frac{E_{\mathrm{fl}}}{2} \sum_{j}^{m} \lambda_j \left(c_j - c_j^{(0)} \right)^2 , \tag{2.55}$$

where the parameters E_{fl} and λ_i of the input file are reinterpreted as force constants. Note, that in order to obtain a restraining harmonic potential, the sign of eq (2.55) has to be inverted by also setting the option `-restrain`.

Multiple Flooding Potentials

More than one flooding potential defined by different sets of eigenvectors $a_{ij}^{(k)}$ and inverted eigenvalues $\lambda_j^{(k)}$ can be applied and combined by concatenation of several `*.edi` files. The input data sets within this concatenated input file are read sequentially and lead to independent sequential execution of the above steps for the respective flooding and/or restraining potentials $V_{\mathrm{fl}}^{(k)}$; the overall flooding force is given by $\mathbf{f}_{\mathrm{fl}} = \sum_k \mathbf{f}_{\mathrm{fl}}^{(k)}$. Therefore, any combination of parameters for adaption and constant flooding as well as different origin structures $\mathbf{c}^{(0)}$ is possible. For a discussion of typical cases where multiple flooding potentials are particularly useful, see end of the next chapter.

Chapter 3

Applications of Flooding

All simulations were carried out using a beta version of GROMACS 3.3, which includes a QM/MM interface [63] to GAUSSIAN03 [185], and the flooding implementation described in the present work. In this chapter, two examples shall illustrate and test the flooding method and its implementation. In the first example, n-butane, the *trans-gauche* conformational transition was accelerated. The second example shows the breakage of a chemical bond in a dialkoxymethylene-cyclopropane in aqueous solution. This chemical reaction is thus predicted in a QM/MM framework.

3.1 *Trans-Gauche* Transition of n-butane

The *trans-gauche* dihedral transition of n-butane (Figure 3.1) is governed by a high energy barrier of 15.9 kJ/mol [186], and thus is unlikely to occur at timescales accessible to MD simulation. This renders n-butane a suitable model system for our illustrative purposes.

Setup

The GROMOS87 united atom force field was used for n-butane [111, 112]. First, the *trans* conformation of n-butane was energy minimized with steepest descent until the maximum force was smaller than $10^{-4}\,\mathrm{kJ(mol\,nm)^{-1}}$. All molecular dynamics simulations, starting from this energy minimized structure, and using the same initial velocities, were carried out with an integration time step of 1 fs and temperature coupling to a 300 K heat bath ($\tau_c = 0.1$ ps) [187].

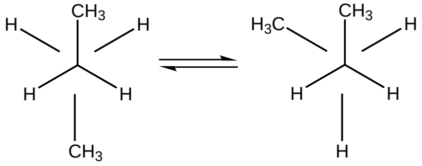


Figure 3.1: *trans-gauche* conformational transition of n-butane.

Two different flooding potentials V_{fl} were generated, the first from PCA, and the second from NMA. For PCA, the covariance matrix **C** was computed from a 100 ps trajectory, during which n-butane stayed in the *trans* conformation, as expected. For NMA, the Hessian matrix **H** was computed from the energy minimized structure. Subsequently, four flooding MD simulations were performed. First, two constant flooding simulations with V_{fl} based on PCA coordinates (MD1) or based on NMA coordinates (MD2), respectively. Second, two adaptive flooding simulations with V_{fl} based on PCA coordinates (MD3) or NMA coordinates (MD4). The flooding subspace was chosen to be spanned by all $m = 6$ internal degrees of freedom of united atom n-butane, i.e., the first six PCA or last six NMA modes, respectively, and the energy minimized structure of the *trans* conformer was used as the reference structure for the rotation / translation correction.

Constant Flooding

To accelerate the *trans-gauche* dihedral transition of n-butane, a flooding potential V_{fl} with constant flooding strength $E_{\text{fl}} = 12.0$ kJ/mol was applied throughout the simulations. Figure 3.2 shows the time evolution of the dihedral angle (A) and the flooding energy V_{fl} (B) for the two simulations MD1 and MD2, based on PCA and on NMA coordinates, respectively. In both simulations, the applied flooding potential induced a *trans-gauche* dihedral transition. As expected for a small molecule like n-butane in vacuo, the PCA and NMA based flooding simulations behave very similarly.

Starting with the *trans* conformation (dihedral = – 180°), the dihedral angle displayed a high motional amplitude and, correspondingly, a low frequency in the broadened minimum (Figure 3.2 A). Within the first few picoseconds of the simulations, the fluctuational amplitude of the dihedral angle was slightly higher in MD2 compared to MD1, resulting in a faster escape to the *gauche* minimum (dihedral = – 60° or −300°), which occurred after 5.9 ps in MD2 compared to 12.4 ps in MD1. The system escaped from the influence of the flooding potential through the dihedral transition. Therefore, V_{fl} dropped to (nearly) zero concomitant with the structural transition.

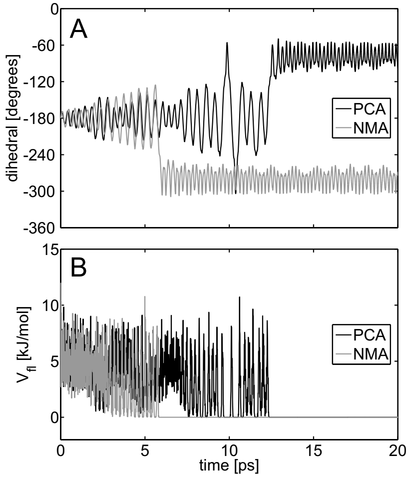


Figure 3.2: Time evolution of the dihedral angle (A) and V_{fl} (B) in the course of the constant flooding simulations MD1 (black) and MD2 (gray).

MD1 and MD2 were started from the same initial configuration, i.e., same structure and velocities. Nonetheless, the slight differences of the applied PCA or NMA based flooding potentials yielded two different *gauche* conformers, which are mirror images and thus equally likely. Before the final transition at 12.4 ps, MD1 transiently visited the *gauche* conformer at 9.8 ps and at 10.3 ps, respectively.

The choice of E_{fl} strongly affects the mean escape time, because the destabilization free energy, ΔF, which determines the acceleration due to the flooding potential, scales with the flooding strength as

$$\Delta F = E_{\mathrm{fl}} \exp\left(-mk_{\mathrm{B}}T/2E_{\mathrm{fl}}\right) \tag{3.1}$$

(eq 33 in Ref. [69]). To demonstrate this, MD2 was repeated with $E_{\mathrm{fl}} = 11.5\,\mathrm{kJ/mol}$ and $12.5\,\mathrm{kJ/mol}$, respectively, starting from identical initial configurations. In the first case, the transition occurred after 19.5 ps, whereas in the latter, the escape time was only 1.5 ps (data not shown), in line with eq 3.1. Therefore, the parameter E_{fl} has to be chosen very carefully. If E_{fl} is chosen too small, one does not achieve a sufficiently large acceleration to observe a transition within the limited simulation time span. In contrast, too strong flooding might alter the free energy landscape also within the crucial transition state region, and might thus lead to artefacts. To allow a more direct control of the expected acceleration, one would wish to adjust the destabilization free energy ΔF directly rather than E_{fl}. This is achieved by the adaptive flooding scheme, in which E_{fl} is updated in every MD time step such as to reach a desired target destabilization free energy ΔF_0.

Adaptive Flooding

To illustrate its benefits, adaptive flooding with a target destabilization free energy $\Delta F_0 = 3\,\mathrm{kJ/mol}$, $E_{\mathrm{fl}}^{(0)} = 0$, and a time constant $\tau_1 = 2\,\mathrm{ps}$ was employed on n-butane. Figure 3.3 shows the time evolutions of the dihedral angle (A) and the flooding energy V_{fl} as well as the flooding strength E_{fl} (B) for the two adaptive flooding simulations MD3 and MD4. Also here, the results of the simulations based on PCA (MD3) and on NMA (MD4) are very similar. The *trans-gauche* transition occurred after 15 ps (MD3) or 59 ps (MD4), respectively, after which the *gauche* conformation was adopted for the rest of the simulation time. Similar to the constant flooding simulations discussed above, the fluctuational amplitude of the dihedral angle within the first few picoseconds of the simulations was slightly higher in the NMA based flooding simulation MD4 compared to the PCA based MD3, resulting in a faster escape to the *gauche* minimum.

Starting at zero, E_{fl} increased at the beginning of the simulations, until it reached a plateau at about $11 \ldots 12\,\mathrm{kJ/mol}$ (Figure 3.3 B). V_{fl} increased accordingly, before it dropped to nearly zero concomitant with the dihedral transition, and should have been switched off at this point. However, for the sake of illustration this was not done here, with the effect that at longer time scales, contrary to constant flooding, the adaptive scheme caused V_{fl} to increase again. After about 50 . . . 60 ps in MD4, the adaptively growing flooding potential started to affect the final *gauche* conformer as well. Hence, to allow for an unperturbed relaxation, flooding should be switched off after the structural transition.

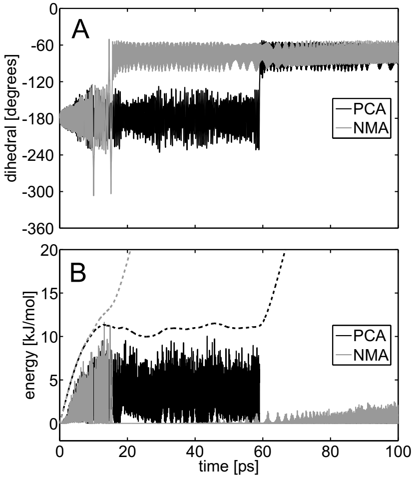


Figure 3.3: Time evolutions of dihedral angles (A) and V_{fl} (B, solid line) as well as E_{fl} (B, dashed line) for MD3 (black) and MD4 (gray).

Figure 3.4 shows the potential energy (with V_{fl} excluded) in the course of the flooding simulation MD3. Here, the flooding potential V_{fl} was switched off manually after the structural transition occurred. To explain Figure 3.4 note that, according to the equipartition theorem, each of the six internal degrees of freedom of n-butane has an average thermal energy of $k_B T/2$, resulting in an average thermal energy of 7.5 kJ/mol at $T = 300\,\mathrm{K}$. Therefore, setting the potential energy of the *trans* minimum to zero, the potential energy of n-butane fluctuated around this value at the start of the simulation. The gradually increasing flooding potential caused a respective increase of the mean potential energy during the first few picoseconds. Upon the dihedral transition, the system escaped the region in conformational space influenced by the flooding potential (cf. Figure 3.3 B) and rapidly relaxed into thermal equilibrium of the *gauche* conformer. Note that the maximum potential energy in Figure 3.4 of ca.

23 kJ/mol surely overestimates the true transition state energy, since the minimum energy path is very unlikely to be followed in a flooding MD simulation at finite temperature. A more reliable estimate of about 12 kJ/mol for the free energy barrier is provided by a sliding gaussian-averaged time dependent partition function (eq 3.3 in Ref. [70]),

$$Z(t) = \int_{-\infty}^{\infty} dt' g_\sigma(t - t') \exp\left(E_{\mathrm{pot}}(t')/k_{\mathrm{B}}T\right), \tag{3.2}$$

see Figure 3.4.

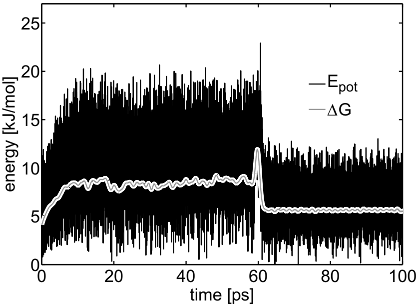
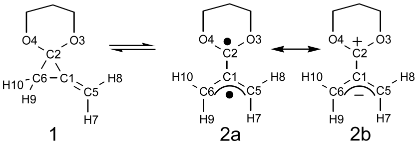


Figure 3.4: Potential energy (black) in the course of MD3. The E_{pot} of the initial structure was set to zero. For the free energy estimate (gray), a width $\sigma = 0.4$ ps was chosen for the Gaussian function g_σ (cf. eq 3.2) .

3.2 MCP

To demonstrate the flooding method for the more challenging case of a chemical reaction in solution, we investigated the rearrangement of the dialkoxy-substituted methylene-cyclopropane (MCP) **1** to the dialkoxy-trimethylenemethane (TMM) intermediate **2** in water, see Figure 3.5. The electronic configuration of TMM **2** can be described as a superposition of the diradicalic structure **2a** and the zwitterion **2b**. The balance between these two configurations is critically influenced by the solvent polarity. Apolar solvents prefer the diradical, whereas increasing solvent polarity stabilizes the zwitterion.

Since molecular mechanical (MM) force fields are unable to describe chemical reactions, we applied a hybrid QM/MM [53] approach. The flooding potential destabilizes the MCP **1**, and triggers the ring-opening reaction towards TMM **2**. Due to the high energy barrier, this bond-breaking is not accessible to unperturbed QM/MM simulations.



1 2a 2b

Figure 3.5: MCP **1** and TMM **2**.

Setup

The QM/MM flooding simulations involved the following steps. First, the system was properly prepared, including the generation of an equilibrated solvent configuration. Then, a free QM/MM equilibration run was carried out to sample data for a PCA. Finally, the QM/MM flooding MD simulation was performed.

Initially, the geometry of MCP was optimized in the gas phase at the HF/3-21G* level and then solvated in a box of 1182 SPC water molecules. Positionally restraining all MCP atoms, the solvent was first energy minimized (steepest descent, 1000 steps) and subsequently equilibrated for 250 ps at 300 K. During these simulations, the whole system was described classically, using the GROMOS96 force field (ffG43a2) for MCP. Application of the SETTLE [188] method allowed for an integration time step of 2 fs. The system was coupled to a 300 K heat bath ($\tau_\mathrm{c} = 0.1\,\mathrm{ps}$) and to a 1-bar pressure bath ($\tau_\mathrm{p} = 1\,\mathrm{ps}$) [187]. For the evaluation of the nonbonded forces, a cutoff of 1.4 nm was applied.

The resulting structure was taken as a starting point for a subsequent 50 ps QM/MM free MD simulation, with MCP treated at the HF/3-21G* level. All QM/MM simulations were carried out without constraints on the QM-QM bonds, and an integration time step of 1 fs was used (all other simulation parameters as described above). In the applied QM/MM scheme [63], the van der Waals interactions between the QM and the MM atoms were described at the force field level, whereas the point charges of the SPC water molecules were included into the QM Hamiltonian and thus polarize the solute wavefunction. For PCA, the covariance matrix **C** of the motion of the eight atoms of the hydrocarbon methylene-cyclopropane ring only (C1, C2, C5, C6, H7 - H10, see Figure 3.5) was computed from the last 40 ps of this trajectory, using the gas phase optimized MCP structure as the reference.

Adaptive flooding with a target destabilization free energy $\Delta F_0 = 75\,\mathrm{kJ/mol}$, a time constant $\tau_1 = 0.1\,\mathrm{ps}$, and $\alpha = 1$ was employed. The flooding subspace was chosen to be spanned by all $3N-6 = 18$ eigenvectors of the constructed covariance matrix $\mathbf{C}$. In the QM/MM flooding simulation, the MCP solute was treated at the UB3LYP/6-31G* level. UB3LYP was chosen, since unrestricted density functional theory, in addition to its capability to accurately describe heterolytic bond cleavage, turns out to work exceptionally well also for diradicalic singlet states resulting from homolytic bond cleavage, at comparably low computational cost [189, 190]. To relax putative strain due to the previously applied Hartree-Fock level, 200 fs free density functional QM/MM molecular dynamics were performed before switching on the flooding potential.

Results

As expected, the applied flooding potential induced the ring-opening reaction from **1** to **2**, as can be seen from Figure 3.6 A, which shows the distance between the C2 and C6 carbon atoms in the course of the flooding simulation. The C2-C6 bond broke, and subsequently, the initially perpendicular C1-C2-C6 and C2-O3-O4 planes became coplanar (Figure 3.6 B). The resulting TMM was then stable for the rest of the simulation time. Through the bond breaking, the system escaped from the influence of the flooding potential, which was switched off manually after 1.5 ps to allow for an unperturbed relaxation (Figure 3.6 C).

In Figure 3.6 D, the potential energy of the solute in the course of the flooding simulation is shown (gray) along with a free energy estimate (black). For comparison, the two different widths of $\sigma = 10\,\mathrm{fs}$ and $\sigma = 20\,\mathrm{fs}$ were chosen for $g_\sigma(t - t')$ in eq 3.2 to obtain the free energy from the potential energy $E_{\mathrm{pot}}(t)$, shown as solid and dashed line, respectively. From the former, an activation free energy barrier of ca. 125 kJ/mol is estimated, whereas the latter yields about 100 kJ/mol. These computed values are close to the experimentally determined activation energy of 107 kJ/mol in C_6D_6 [191]. As already stated above for n-butane, the minimum energy pathway can not be expected to be followed in a flooding simulation, and therefore, the obtained energy barrier represents an upper bound. In contrast to the unsubstituted trimethylenemethane in the gas phase [70], the TMM **2** is a stable intermediate in water and persists throughout the rest of the simulation time. The free energy of **2** is not only significantly lower than that of the transition state, but, after full relaxation, even lower than that of MCP **1** (Figure 3.6 D). The stability of the resulting TMM **2** compared to the unsubstituted trimethylenemethane in the gas phase due to the polar water solvent and the dialkoxy substituents clearly advocates the formation of the zwitterion **2b**, which is also underlined by the distribution of Mulliken atomic partial charges (-0.63 for the allylic moiety (atoms C1, C5, C6, H7 - H10) and +0.63 for the remaining atoms, values computed for snapshot after 2 ps).

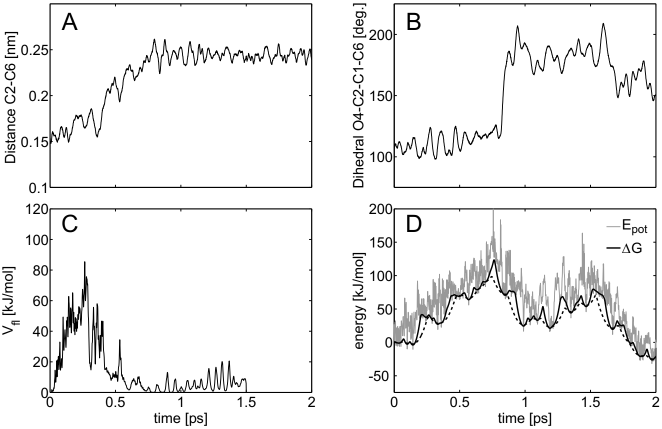


Figure 3.6: Time evolutions of the C2-C6 bond breakage (A), the dihedral angle between the C1-C2-C6 and C2-O3-O4 planes (B), V_{fl} (C), and potential (gray, D) as well as free energy estimate (black, D) in the course of the adaptive flooding simulation.

3.3 General Remarks

In the following, some aspects will be discussed which go beyond the presented examples and might provide some additional help to the user to successfully apply the flooding method. From our experience, suitable flooding potentials V_{fl} can be generated from relatively short, i.e., $1 \ldots 10$ ns free molecular dynamics trajectories of the initial educt conformation. This is backed up by a recent study, which has shown that for the small protein crambin, convergence of a medium sized ($m = 100$) conformational subspace is remarkably fast [192]. In particular, the *directions* of conformational motion in the conformational subspace are already correctly described by a 5 ns MD trajectory. Furthermore, the method does not require fully converged ensembles. Accordingly, in our flooding implementation, the fluctuation *amplitudes*, i.e., eigenvalues of the covariance matrix, which need much longer to converge [193, 194], are adjustable through an empirical parameter α. From our experience, scaling the fluctuation amplitudes obtained from short MD simulations by $\alpha^2 = 2$ is a good choice for proteins. For small molecules, which can be sampled sufficiently long such that also the fluctuation amplitudes are converged, scaling is not required.

Multiple flooding potentials are useful for the following cases (see Sec. 2.6.2). First, to examine multi step pathways [195]. To this end, the number of applied flooding potentials can be increased successively, i.e., after the escape from the initial conformation and proper relaxation of the accessed product state, a second flooding potential to additionally destabilize this new state might be added to the first one, etc. This procedure is required, since, by construction, V_{fl} acts on the energy landscape only in the vicinity of the initial structure. To find alternative transitions and to avoid revisiting already known (and therefore uninteresting) conformations, additional flooding potentials to destabilize the known states can be applied simultaneously. Second, transitions from *anharmonic* minima can be accelerated by multiple flooding potentials, appropriately centered at different origin structures $\mathbf{c}^{(0)}$. Third, degenerate stereochemical rearrangements of the educt (i.e., leading to enantiomers of the educt) can be obviated by simultaneous application of multiple flooding potentials, which can be generated using symmetry.

We have described both, constant flooding as well as adaptive flooding. Which of these two techniques shall be applied for a particular case? In a constant flooding simulation, the system propagates on a stationary free energy landscape, and properly chosen parameters accelerate spontaneous transitions after an initial equilibration phase. In contrast, adaptive flooding implies a time-dependent free energy landscape. The advantage of the latter approach is that a transition is guaranteed to occur within a controllable simulation time, as long as a sufficiently large target destabilization free energy ΔF_0 and small flooding time constant τ are chosen. On the other hand, if ΔF_0 is too large and/or τ is too small (tight coupling), too little time is provided to sufficiently equilibrate in the rapidly changing free energy landscape. This might lead to artefacts, i.e., the system might be driven towards high energy configurations before an exit channel from the altered region of the free energy landscape is found. Thereby, artificially large activation energies and possibly wrong pathways and product states would be predicted. Slow degrees of freedom, e.g., in proteins, require long equilibration times [196]. Therefore, to induce conformational transitions in proteins, adaptive flooding with sufficiently small ΔF_0 and large τ or constant flooding should be applied. On the other hand, for systems displaying fast degrees of freedom only, e.g., the MCP example discussed above, adaptive flooding with tight coupling is well suited and the technique of choice.

In both cases, the flooding strength E_{fl} critically determines the mean escape time (see Sec. 3.1). The proper choice of the parameter E_{fl} is therefore very important, since too weak flooding potentials will not sufficiently accelerate conformational transitions, whereas too strong flooding potentials heavily distort the energy landscape and lead to artefacts. However, in the majority of cases, the heights of the energy wells which have to be surmounted are unknown. We therefore suggest to first apply adaptive flooding with large destabilization free energy for an initial exploration of the energy landscape, and to extract appropriate parameters from this for subsequent more gentle flooding simulations, which will then yield improved reaction pathways. From our experience, to accelerate conformational transitions in proteins by means of a constant flooding potential, flooding strengths of $0.2 \ldots 3\, k_{\mathrm{B}}T$ per degree of freedom influenced by the flooding potential are suitable.

Chapter 4

The Fluoroprotein asFP595: Spectra and Proton Paths

Fluorescence is widely used by marine organisms for visual communication (e.g., to attract prey) or for protection against UV damage [197, 198]. The discovery of the green fluorescent protein (GFP) from the jellyfish in the early 1960s ultimately heralded a new era in cell biology. GFP can be tagged to other proteins in living cells, and the readout of its fluorescence emission with an optical microscope enables to monitor cellular processes in living organisms, such as protein expression, protein-protein interactions, and protein trafficking [72, 73, 74]. However, optical microscopy based on GFP fluorescence is usually limited due to diffraction and photobleaching [77]. To overcome these drawbacks requires a new generation of proteins that can be can be reversibly photoswitched between a fluorescent *(on)* and a non-fluorescent *(off)* state [6, 7, 8, 199, 200, 201, 202]. The GFP-like fluoroprotein asFP595, isolated from the sea anemone *Anemonia sulcata*, is a prototype for such a reversibly switchable fluorescent protein (RSFP). RSFPs can lead to increased resolution in optical microscopy, since fluorescence emission from regions that are out of the focus of the microscope can be selectively (and reversibly) depleted. Thereby, the size of the focal spot is squeezed to below 50 nm [77]. Likewise, RSFPs will enable the repeated tracking of protein movement in single cells [203]. Since fluorescence can be sensitively read out from a bulky crystal, the prospect of erasable three-dimensional data storage is equally intriguing [80].

The work presented in Chapters 4 and 5 of this thesis was done in close collaboration with Dr. Gerrit Groenhof. Furthermore, a close collaboration with Dr. Martial Boggio-Pasqua and Prof. Mike Robb from the Imperial College London was established. All static multiconfigurational calculations in the gas phase were done by Martial. I carried out all MD simulations, analyzed the data, and prepared the figures.

4.1 Structure and Spectroscopic Properties

The asFP595 protein is structurally and spectroscopically well characterized [8, 7, 204, 205], but its detailed photoswitching mechanism at the atomistic scale remains largely unknown. In particular, it is unclear to which extent different protonation states of the chromophore and associated proton transfer events between the chromophore and the protein surrounding determine the optical properties of the protein and, if they do, what their pathways are. The asFP595 protein can be switched from its non-fluorescent *off* state to the fluorescent *on* state by irradiation with green light of wavelength 568 nm (2.18 eV). From this so-called kindled *on* state, the same green light elicits a red fluorescence emission at 595 nm. Upon kindling, the absorption maximum at 2.18 eV is diminished, and an absorption peak at 445 nm (2.79 eV) appears. This peak at 2.79 eV was tentatively attributed to the absorption of a different protonation state [204]. However, the involved protonation states were unknown.

The kindled *on* state can be promptly switched back to the initial *off* state by blue light of wavelength 445 nm. Alternatively, the *off* state is repopulated through thermal relaxation within seconds. In addition, if irradiated with intense green light over a long period of time, asFP595 can also be irreversibly converted into a fluorescent state that cannot be quenched by light [7]. The nature of this state is hitherto unknown.

The switching cycle of asFP595 is reversible and can be repeated many times without significant photobleaching. These properties render asFP595 a promising fluorescence marker for high-resolution optical far-field microscopy, as recently demonstrated by Hell and coworkers [79]. Currently, however, with its low quantum yield ($< 0.1\%$ and 7% before and after activation, respectively [8, 203]) and rather slow switching kinetics, the photochromic properties of asFP595 need to be improved. To systematically exploit the potential of such switchable proteins and to enable rational improvements to the properties of asFP595, a detailed molecular understanding of the photoswitching mechanism is mandatory.

High-resolution crystal structures of wild-type (wt) asFP595 in its *off* state [205, 206, 207], of the Ser158Val mutant in its *on* state, and of the Ala143Ser mutant in its *on* and *off* states [205] were recently determined. Similar to GFP, asFP595 adopts a β-barrel fold enclosing the chromophore, a 2-acetyl-5-(p-hydroxybenzylidene)imidazolinone (Figure 4.1). The chromophore is post-translationally formed in an autocatalytic cyclization-oxidation reaction of the Met63-Tyr64-Gly65 (MYG) triad. As compared to the GFP chromophore, the π-system of MYG is elongated by an additional carbonyl group [208]. Reversible photoswitching of asFP595 was possible even in whole protein crystals, and x-ray analysis showed that the *off-on* switching of the fluorescence is accompanied by a conformational *trans-cis* isomerization of the chromophore [205]. The observed absorption blue-shift from 2.18 eV to 2.79 eV upon kindling suggests that the photoswitching of asFP595 is also accompanied by protonation state changes of the chromophore [204]. Additional evidence comes from the related GFP, where proton transfer processes dominate photoswitching. In GFP, an anionic and a neutral chromophore protonation state are interconverted through a proton relay mechanism [147, 148, 149, 209, 210]. Because protons were not observed in the crystal structures, their role and possible transfers during kindling of asFP595 remained largely unclear. The aim of the present work is to investigate the detailed molecular mechanism by which these two conformers are interconverted, as well as to study associated proton transfer events.

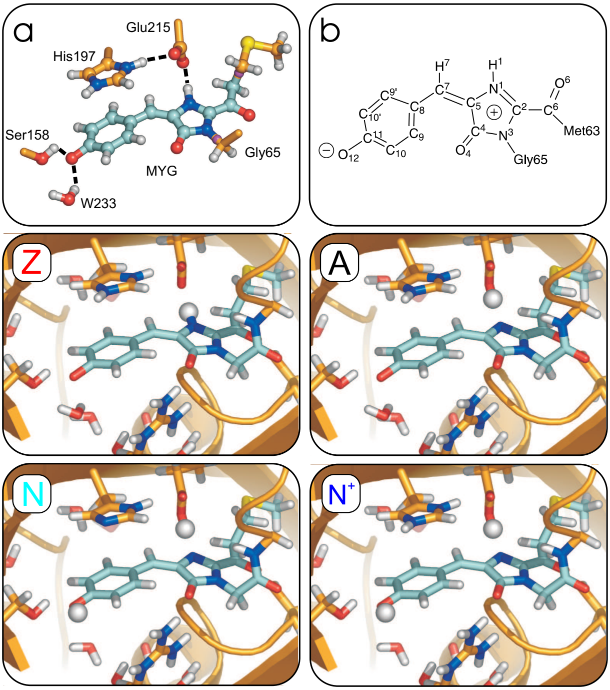


Figure 4.1: Zwitterionic asFP595 chromophore (MYG) in the dark state *trans* conformation. **a)** MYG and adjacent side chains of binding pocket residues. Glu215, Ser158, and the crystallographic water molecule W233 are hydrogen bonded to MYG, and His197 is π-stacked to the MYG phenolate. The carbon skeleton of the QM subsystem used in the QM/MM calculations (see Section 4.3.2) is shown in cyan, the MM carbon atoms are shown in orange, and the hydrogen link atoms capping the QM subsystem are shown in magenta. **b)** Schematic drawing of MYG, defining the atom names used in the text. The protonation states of the chromophore pocket in wild-type asFP595 considered in the present work are a zwitterionic state (Z), an anionic state (A), and a neutral state, with His197 singly (N) or doubly protonated (N+). Hydrogen atoms characterizing the protonation states are highlighted as spheres.

4.2 Proton Paths and Absorption Spectra

To investigate the role of different protonation states and proton translocations, optical absorption (UV/Vis) spectra were calculated from MD ensembles of a series of asFP595 protonation states, shown in Figure 4.1. To this end, the semi-empirical INDO/S (ZINDO) method [120, 211] and time-dependent density functional theory (TDDFT) [126, 212, 127, 213, 128] were employed. In particular, TDDFT has been used to predict absorption energies for a wide variety of biological chromophores, including GFP [214, 215, 216, 217, 218]. Recently, single-structure TDDFT calculations on a small subsystem of asFP595 have been attempted [219].

Since the spectra of biological chromophores can be strongly tuned by their protein environment, the TDDFT calculations were performed within the QM/MM approach [53]. The chromophore was treated quantum-mechanically, whereas the rest of the system was described at the force field level (Figure 4.1a). To properly account for the protein environment in the INDO/S calculations, the majority of the chromophore binding pocket was treated at the QM level (see Section 4.3.2). Comparison of the calculated UV/Vis spectra to the available experimental data, in conjunction with computed protonation probabilities of the titratable groups in the chromophore cavity, would allow us (i) to unambiguously assign the protonation patterns of the *off* and *on* states of wt asFP595, and (ii) to provide atomistic insights into the proton transfers that interconvert the protonation states involved in photoswitching.

4.3 Simulation Details

The UV/Vis spectra were calculated in three steps. First, extensive classical MD simulations for different protonation states of wild-type asFP595 (i.e., with the *trans* chromophore) and of a *cis* mutant structure were performed to generate thermodynamic ensembles. Second, for each protonation state, a sufficient number (100) of structures were extracted from these ensembles and modestly relaxed. Finally, the excitation energies were calculated for these relaxed structures, and superimposed. In the QM/MM TDDFT calculations, the chromophore was described quantum mechanically at the TD-BVP86/6-31G* level of theory [220, 221, 222], and the rest of the protein, water, and ions were described at the force field level (Figure 4.1a). In the INDO/S calculations, the majority of the chromophore binding pocket, composed of the entire MYG chromophore and the side chains of residues Lys67, Arg92, Glu145, Ser158, His197, Glu215, and the crystallographic water molecule W233 were described at the quantum level.

4.3.1 MD Simulations

To generate thermodynamic ensembles, force field based MD simulations of monomeric wild type (wt) asFP595 (*trans* chromophore) and of a *cis* mutant structure were carried out for five different protonation states of the chromophore pocket. These five protonation states covered all conceivable chromophore protonation states: the zwitterionic state “Z”, the anionic state “A”, a neutral state “N”, a neutral state “N+” (where + indicates a doubly protonated His197), and a double-anionic state “D”. A doubly protonated cationic chromophore was not considered due to its high acidity [223]. After minimization and equilibration, each system was simulated for 7.5 ns. The starting coordinates for the simulations were taken from the 1.3 Å x-ray crystal structure (PDB entry 2A50, Ref. [205]). All simulations were performed using the Gromacs simulation package [183] together with the OPLS all-atom force field [108].

The MD simulations were performed in a rectangular periodic box of about $8 \times 8 \times 8\,\mathrm{nm}^3$. Each system contained in total about 15,000 TIP4P water molecules, including 340 crystallographic water molecules. After assigning the protonation pattern of the chromophore pocket, all other polar, aromatic, and aliphatic hydrogens were added to the protein with the HB2MAK [224] routine of WHATIF [225]. To each of the systems, sodium and chloride ions at physiological concentration were added to compensate for the net charge of the protein. The actual number of ions used depended on the total charge of the protein, which differed for the five chosen protonation patterns of the chromophore cavity. The final systems contained around 68,000 atoms. Prior to the simulations, the systems were energy minimized for 1000 steps using steepest descent. Subsequently, a 200 ps MD simulation was performed with harmonic position restraints on all heavy protein atoms (force constant $1000\,\mathrm{kJ\,mol^{-1}nm^{-2}}$) to equilibrate the water and the ions.

Simulations were run at constant temperature and pressure by coupling to external baths [187] ($\tau_\mathrm{T} = 0.1\,\mathrm{ps}$, $\tau_\mathrm{p} = 1\,\mathrm{ps}$). LINCS [226] was used to constrain bond lengths, thus allowing a time step of 2 fs for the force field simulations. SETTLE [188] was applied to constrain the internal degrees of freedom of the water molecules. A twin-range cut-off method was used for the Lennard-Jones interactions. Interactions within

1.0 nm were updated every time step, whereas Lennard-Jones interactions between 1.0 and 1.6 nm were updated every ten steps. Coulomb interactions within 1.0 nm were computed each step as well. Beyond this cut-off, the particle mesh Ewald (PME) method [227] with a grid spacing of 0.12 nm was used.

To obtain force field parameters for MYG, quantum chemical calculations were carried out for the isolated chromophore with Gaussian03 [185]. The model chromophore used for these calculations was capped with methyl groups. The atomic partial charges for the ground state (optimized at the B3LYP/6-31+G* level) were estimated by fitting to the molecular electrostatic potential of the chromophore in vacuo according to the CHELPG scheme [228], both for the *cis* and *trans* conformations (see Appendix). The partial charges for the excited state were obtained from a configuration interaction calculation including singly excited Slater determinants (CIS/6-31+G*). Heavy-atom bond lengths and angles were taken from the crystal structure, those involving hydrogen atoms from the QM-optimized structures. For the force constants and Lennard-Jones parameters, OPLS force field parameters of residues with similar chemical nature were adopted (His, Tyr, Met, styrene). Except for partial charges, the same parameters were used for the *trans* and *cis* conformations.

4.3.2 UV/Vis Spectra

The optical absorption spectra of asFP595 were calculated for ensembles comprised of a sufficiently large number of structures (100 structures taken at equally-spaced time frames from each force field trajectory). The structures were modestly relaxed by means of very short QM/MM geometry optimizations (50 steps steepest descent, HF/3-21G*//OPLS, ONIOM) [154]. The value of 50 minimization steps was small enough to leave the thermodynamic ensemble nearly unperturbed, yielding improved configurations for the calculation of the excitation spectra. Finally, UV/Vis spectra were composed from the calculated absorption energies ΔE_{ji}^{max} via superposition of Gaussian functions,

$$G\left(\Delta E\right) = \sum_{i=1}^{100} \sum_{j=1}^{3} f_{ji} \exp\left[-\frac{\left(\Delta E - \Delta E_{ji}^{max}\right)^2}{2\sigma^2} \right].$$

Here, the ΔE_{ji}^{max} are the excitation energies of the first three excited singlet states ($j = 1, 2, 3$) of structure i, and f_{ji} is the corresponding oscillator strength. A width of $\sigma = 0.02\,\mathrm{eV}$ was chosen.

In the QM/MM TDDFT calculations, the BVP86 density functional was used to describe the chromophore. This density functional method has been applied to calculate excitation energies for various systems ranging from atoms and small molecules [229] to large bio-organic systems like the free base porphin [230]. In addition, a recent TDDFT study comparing a number of density functionals showed that this functional yields the best excitation energies for the GFP chromophore in the gas phase [231]. To estimate convergence with respect to the basis set size, the excitation energies of representative structures taken from each of the MD ensembles were re-evaluated using the more complete 6-31+G* and 6-311G* basis sets. The remainder of the system, consisting of the apo-protein, water molecules, and ions was modeled with the OPLS force field. The $C_\beta - C_\gamma$ bond of the Met63 side chain and the $N - C_\alpha$ bond of the Gly65 backbone connecting the QM and MM subsystems were replaced by constraints, and the QM part was capped with hydrogen link atoms. The force on the link atom was distributed over both atoms of the respective bonds according to the lever rule. The QM subsystem experienced the Coulomb field of all MM atoms within a sphere of 1.6 nm radius, and Lennard-Jones interactions between QM and MM atoms were also included. To avoid over-polarization of the QM subsystem by the MM charges, the partial charges of the MM atoms next to the link atoms as well as of the hydrogen atoms bound to these atoms were set to zero. The charge of the sulphur atom in the Met63 side chain was scaled to maintain an integer charge on the MM subsystem. In addition, all charges in the MM subsystem were described by Gaussian charge distributions during the TDDFT calculations. These Gaussian distributions were centered on the MM atom and had a width of $\sigma = 0.3\,\mathrm{nm}$ [156].

In the INDO/S calculations, a large part of the chromophore cavity, composed of the entire MYG chromophore and the side chains of residues Lys67, Arg92, Glu145, Ser158, His197, Glu215, and the crystallographic water molecule W233 were described at the quantum level. To this end, the following bonds were cut and the open valences were capped with hydrogen atoms: the $C_\gamma - C_\beta$ bonds of residues Glu145, His197, and Glu215; the $C_\delta - C_\gamma$ bond of Arg92, the $C_\beta - C_\alpha$ bond of Ser158; the $C_\epsilon - C_\delta$ bond of Lys67; and the $N - C_\alpha$ bond of the Gly65 backbone. All calculations were performed with Gromacs 3.3 [183] and its QM/MM interface [63] to Gaussian03 [185]. Modifications were made in the one-electron integral routines of Gaussian03 for the TDDFT computations of the chromophore polarized by the Gaussian charge distributions of the MM atoms.

4.3.3 Poisson-Boltzmann Electrostatics

To quantify the population of the different protonation states of the chromophore pocket, Poisson-Boltzmann electrostatics (PBE) calculations [232] were performed on wt-asFP595 (pdb entry 2A50 [205]) and a mutant structure. The x-ray structure of the mutant shows density for both *cis* and *trans* chromophores, thus allowing the effect of *trans-cis* photoisomerization on the protonation state populations to be evaluated. This work was done in close collaboration with Dr. Astrid Klingen from the group of Prof. Matthias Ullmann, University of Bayreuth. The actual PBE calculations were done by Astrid, whereas I carried out the minimizations of the protein structures and the DFT calculations of the reference pK values in water.

To prepare the crystal structures for the PBE calculations, a monomer was extracted from each structure, and hydrogen atoms were added with WHATIF [224, 225]. Subsequently, the positions of all hydrogen atoms and of the chromophore atoms were optimized using GROMACS and the OPLS force field (l-bfgs, 200 steps). Protonation probabilities were then calculated for all protonatable sites in the protein using the multiflex routine of the MEAD program package [233] and a Metropolis Monte Carlo (MC) algorithm to sample the protonation state energies [234].

In the PBE calculations, partial charges from the CHARMM27 parameter set [110] were used for the protein. Partial charges for the chromophore were derived as described above. Bondi radii [235] were used except for hydrogen $\left(R_{\mathrm{H}} = 1.0\,\text{Å}\right)$. The dielectric constants were set to $\epsilon = 4$ for the protein and $\epsilon = 80$ for the solvent, the ionic strength to $I = 0.1\,\mathrm{M}$, and the temperature to 300 K. For characterizing the titration behavior of the different sites in the protein, pK_{A}-values of appropriate model compounds in aqueous solution were used as reference values. Standard model compound pK_{A}-values [232] were used for the sidechains of Asp, Glu, Lys, Arg, Tyr and Cys. Histidines were treated as described by Bashford *et al.* [236] The pK_{A}-value of the chromophore imidazolinone (N1, Figure 4.1B) was estimated according to the thermodynamic cycle 1 described in Ref. [237] at the B3LYP/6-31G* level of theory using the polarizable continuum water model [238, 239, 240, 241, 242]. The calculated values strongly depend on the chromophore conformation and are 9.1 and 4.7 for *trans* and cis, respectively. Because the pK_{A} of the chromophore phenolate group (O12, Figure 4.1B) does not critically depend on the conformation, we used the value of 8.2 measured for the structurally similar GFP chromophore [223].

The MC calculations involved 500 equilibration scans and 20,000 production scans for every pH-step [243]. Each MC scan comprised N steps, with N being the total number of protonatable sites. In each MC step, the protonation state of one group was changed, and the change in energy was evaluated. In double (triple) MC steps, the protonation state of two (three) groups was changed simultaneously. Double (triple) MC steps were applied to groups with an interaction energy larger than 2 (3) pK-units.

From the calculated protonation probabilities at a given pH, a subset of N residues likely to change their protonation state were identified in each structure: these residues have a protonation probability $\langle x \rangle(\mathrm{pH}) \in [0.1, 0.9]$. For residues in these subsets, we have calculated the relative energies of all 2^N possible protonation states directly from the MEAD output [232]. The protonation forms of all other residues were kept fixed (residues with $\langle x \rangle(\mathrm{pH}) < 0.1$ were considered deprotonated, residues with $\langle x \rangle(\mathrm{pH}) > 0.9$ were considered protonated). The chromophore and the potentially relevant residues Glu215 and His197 were always included into the subset of N residues for which all protonation state energies were computed. The probability of state n was calculated from its relative energy g_n according to

$$p_n(\mathrm{pH}) = \frac{\exp\left(-g_n(\mathrm{pH})/RT\right)}{\sum_n^{2^N} \left[\exp\left(-g_n(\mathrm{pH})/RT\right)\right]}.$$

4.4 Results and Discussion

4.4.1 Protonation States from First-Principles UV/Vis Spectra

Figure 4.2 shows the UV/Vis spectra for the considered chromophore protonation states of asFP595, defined in Figure 4.1. The spectra shown in panels a–e were calculated at the semi-empirical INDO/S level, and those shown in panels e,f at the TDDFT level. Examination of the molecular orbitals revealed that, for all cases, the dominant excitations are $\pi \rightarrow \pi^*$ transitions.

The UV/Vis spectra computed at the INDO/S level are close to the measured spectra and allow us to assign the different protonation states

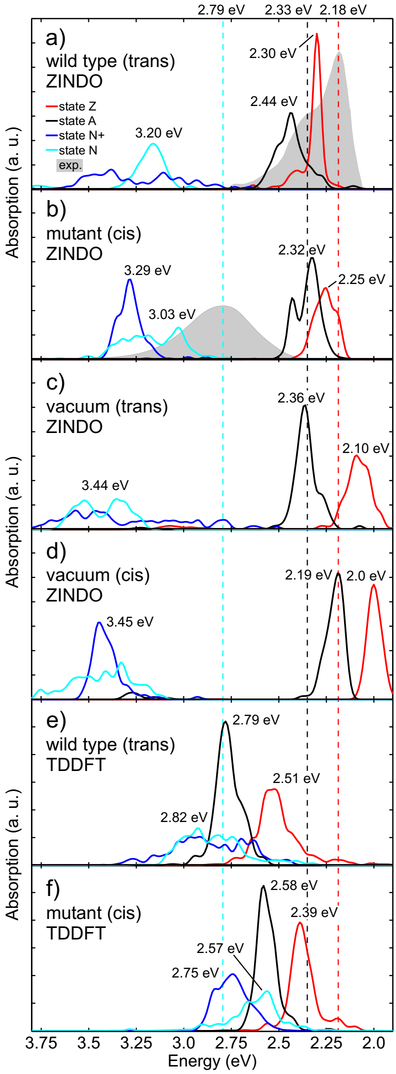


Figure 4.2: Calculated optical absorption spectra of the asFP595 protonation states shown in Figure 4.1, calculated at the semi-empirical INDO/S (panels a–d) and TDDFT levels of theory (panels e, f). **a)** Spectra for the wt protein (*trans* chromophore). **b)** Spectra for the *cis* mutant. The measured absorption spectrum is shown in gray with the maxima indicated by dashed lines. Upon kindling, the zwitterionic and anionic absorptions at 2.18 eV and at 2.33 eV (dashed red and black lines, respectively) are depleted, concomitant with a rising peak at 2.79 eV (dashed cyan line). **c,d)** INDO/S spectra of the chromophore in vacuo, calculated from the same ensemble of structures as in panels a and b. **e,f)** QM/MM TD-BVP86 spectra. The experimental spectrum was adopted from ref [204].

to the measured absorption bands. The TDDFT absorption spectra show the same order of states as the INDO/S spectra, although the absorption energies of the anionic and the zwitterionic chromophores are blue-shifted too much, as discussed in detail below. For wt asFP595, i.e., with the chromophore in the *trans* conformation, the zwitterion (state Z, red curve) displays the most red-shifted absorption, with a strong and narrow maximum at 2.30 eV and a distinct shoulder at 2.41 eV (Figure 4.2a). Transfer of the proton from the imidazolinone nitrogen (N1, Figure 4.1b) to the Glu215 side chain results in the anionic state A, for which we find a strong absorption band at 2.44 eV, with a shoulder at 2.51 eV (black curve). Further protonation of the phenolate oxygen (O12, Figure 4.1b) leads to the N+ state and causes a significant blue-shift towards 3.2 eV (blue curve); this absorption band is weaker and much broader than those of the ionic species. The neutral chromophore state N, with singly protonated His197, also absorbs at around 3.2 eV (cyan). A doubly anionic state with a negative chromophore and a deprotonated Glu215 was not considered, because the close contact between the chromophore and the Glu215 side chain observed in the crystal structure (2.7 Å) requires a hydrogen bond, as shown in Section 7.3.4.

To reveal the effect of *trans*-*cis* isomerization on the absorption characteristics of asFP595, we calculated the spectra not only for the wild type protein, but also for a *cis* mutant (Figure 4.2b). For all protonation states, the absorption of the *cis* isomer is slightly red-shifted by about 0.05 eV to 0.2 eV with respect to that of the *trans* isomer, a result that is in agreement with recent experiments [9]. Therefore, the *trans*-*cis* isomerization alone — without concomitant protonation state changes — cannot explain the observed blue-shift from 2.18 eV to 2.79 eV upon kindling.

As the chromophore of the wt adopts the *trans* conformation in the crystal structure [205, 206, 207], a comparison of the calculated absorption spectra of the *trans* wt (Figure 4.2a) with the experimental spectrum allows us to assign the protonation state of the chromophore cavity in wt asFP595. Under physiological conditions, asFP595 has a measured sharp absorption maximum at 2.18 eV, and a shoulder which is blue-shifted by 0.15 eV at 2.33 eV (gray curve in Figure 4.2a) [204]. Taken together, the calculated sharp maximum of the zwitterion at 2.30 eV and the weaker and broader band of the anion at 2.44 eV yields a spectrum that is only slightly blue-shifted with respect to the measured one (by 0.12 eV). In particular, the oscillator strengths and peak shapes are in good agreement with the experimental spectrum, and the calculated energy difference of 0.14 eV between the zwitterionic and anionic absorption maxima is very accurate. These results suggest that the zwitterion is the dominant species under physiological conditions and that the anion leads to the observed shoulder in the absorption spectrum. Additional support for this assignment comes from the measured absorption spectrum of the anionic asFP595 chromophore in aqueous solution, which has a maximum at 2.38 eV [208].

In both the *trans* and the *cis* conformations, the neutral chromophore absorption at 3.20 eV (*trans*) and 3.03 eV (*cis*) is considerably blue-shifted compared to that of the ionic species. Upon kindling, a rising peak at 2.79 eV is observed (gray curve in Figure 4.2b), which within the estimated accuracy of INDO/S (see below) can be assigned to a population of the neutral *cis* species. The TDDFT spectra support this assignment, see Figure 4.2e and f. Our results strongly suggest that the *trans*-*cis* isomerization of the chromophore is accompanied by protonation state changes and that the absorbing species populated upon kindling is the neutral *cis* chromophore.

To elucidate the electrostatic and steric effects of the protein environment on the absorption characteristics, we studied the absorption characteristics in vacuo, both for the MD ensembles and for optimized structures. Both effects slightly blue-shift the absorption energies of the zwitterion and the anion as compared to the gas phase. Figure 4.2c,d show the spectra of the asFP595 chromophore in vacuo. The electrostatic influence of the protein matrix blue-shifts the absorption of the zwitterion and of the anion by about 0.1 eV to 0.2 eV, whereas the neutral states are slightly red-shifted. In addition to this electrostatic effect, the non-planarity of the chromophore imposed by the protein matrix also leads to a blue-shift in absorption, irrespective of the protonation state (Table 4.1). This latter effect is especially pronounced for the neutral chromophore. However, these shifts do not change the order of the absorption bands.

The degree of chromophore planarity in asFP595 is quantified in Figure 4.3, which shows the distributions of the two ring-bridging dihedral angles of the chromophore in the ensembles used for the excitation calculations. The distributions are independent from the protonation state, but depend on the chromophore conformation. In the *trans* conformation, the protein matrix enforces the hydroxyphenyl ring to deviate from planarity by as much as 20° (Figure 4.3a). In contrast, the *cis* chromophore is considerably more planar (Figure 4.3b), explaining the red-shift. The imidazolinone moiety slightly deviates from planarity by about 5–10°, irrespective of the chromophore conformation (Figure 4.3c and d). The result that the protein environment leads to a *cis* chromophore that is more planar than the *trans* chromphore supports the hypothesis that ultra-fast radiationless deactivation of the dark *trans* state is facilitated by the non-planarity of the chromophore [9].

A comparison of the UV/Vis spectra calculated at the INDO/S level (Figure 4.2a–d) to the spectra obtained at the TDDFT QM/MM level (panels e,f) reveals that the TDDFT absorption bands of the zwitterion and the anion are too much blue-shifted. In fact, none of the absorption bands is close to the measured absorption at 2.18 eV. As shown in Table 4.1, the use of larger basis sets in the TDDFT calculations did not substantially improve the results.

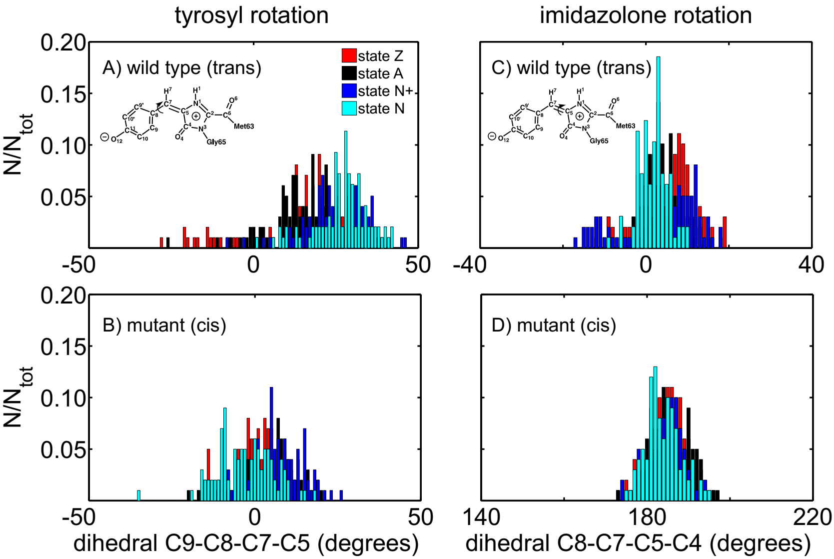


Figure 4.3: Histogram of the ring-bridging dihedral angles of the asPF595 chromophore in the protein matrix. (A) and (B) show the C9-C8-C7-C5 dihedral angle in the *trans* wild type and in the *cis* mutant, respectively. (C) and (D) show the respective C8-C7-C5-C4 dihedral angle distributions. See Figure 4.1b for atom numbering.

However, the absorption band of the neutral species is described more

	TD-BVP86						ZINDO					
state	cis (mutant)			trans (wild type)			cis (mutant)			trans (wild type)		
	6-31G*	6-311G*	6-31+G*	6-31G*	6-311G*	6-31+G*	non-planar protein	non-planar vacuum	planar vacuum	non-planar protein	non-planar vacuum	planar vacuum
Z	2.39	2.38	2.34	2.51	2.51	2.46	2.25	2.00	1.90	2.30	2.10	1.98
N	2.62	2.60	2.58	2.82	2.82	2.79	3.03	3.44	3.00	3.20	3.44	2.96
N+	2.74	2.73	2.70	2.82	2.82	2.80	3.29	3.45	3.00	3.20	3.44	2.96
A	2.58	2.56	2.56	2.79	2.77	2.75	2.32	2.19	2.16	2.44	2.36	2.20

Table 4.1: Calculated absorption energies in eV of different asFP595 protonation states. To elucidate the basis set dependence of the TDDFT absorption energies, representative structures were taken from the maxima of the respective absorption peaks in Figure 4.2. The planar species were optimized in vacuo prior to the excitation calculation.

accurately at the TDDFT level as compared to INDO/S. Furthermore, both methods predict the same relative order of the absorption bands of the different protonation states (Z, A, N, N+), a result that provides further support for our protonation state assignment. The reason for the failure of TDDFT for the zwitterion and the anion is that the $S_0 \rightarrow S_1$ transitions involve intramolecular charge transfer (CT) from the phenoxy- to the imidazolinone-moiety. Such CT excitations are known to be usually rather poorly described by TDDFT [244, 129]. We performed RASSCF(18,7+4+5)[2,2]/6-31G* calculations to further analyze the failure of TDDFT for the ionic protonation states. These calculations revealed that the $S_1(\pi, \pi*)$ state is a strongly doubly-excited state, explaining the too much blue-shifted TDDFT excitation energy.

Controls

We calculated the absorption energies of an anionic and a neutral protonation state of a GFP model chromophore, which is very similar to MYG, to assess the accuracy of the applied methods for asFP595. Since in GFP, the dependence of the chromophoric absorption on the surrounding amino acids is weak [245], we used an isolated model chromophore. INDO/S predicts absorption energies of 2.65 eV and 3.65 eV for the geometry-optimized (HF/6-31G*) anionic and neutral chromophores, respectively, whereas TD-BVP86/6-31G* predicts 3.12 eV and 3.39 eV, respectively. The measured absorption maxima of the anionic and neutral GFP are 2.59 eV and 3.12 eV, respectively [209]. Thus, the red-shifted absorption of the ionic species is well described at the INDO/S level, whereas the excitation energy of the neutral species is overestimated by about 0.5 eV. In contrast, TDDFT is more accurate for the neutral species but fails to predict the correct absorption of the anion with an error of about 0.5 eV. As INDO/S apparently performs better than TDDFT for the red-shifted absorption regime of asFP595 (around 2.18 eV), we based our protonation state assignment in this regime on the INDO/S spectra only. For the more blue-shifted regime (around 2.78 eV), both INDO/S and TDDFT spectra were taken into account.

The MD simulation of the doubly anionic state allows to test our assumption that the close contact between the chromophore imidazolinone ring and Glu215 under physiological conditions requires a hydrogen bond, and, hence, the doubly anionic chromophore needs not be considered. Already at the start of the 7.5 ns force field MD simulation of this protonation state, the distance between the imidazolinone moiety and Glu215 quickly increased from the x-ray value of 2.7 Å to about 4 Å, a result that rules out this protonation state under physiological conditions.

To check whether and how the description of the π-stacking of the His197 imidazole ring and the chromophore affects the calculated excitation energies, we performed an additional QM/MM TDDFT calculation in which we included the imidazole side chain of His197 into the QM subsystem. A structure from the zwitterion ensemble was taken for this purpose. According to this calculation, inclusion of the His197 imidazole ring did not alter the nature of the $\pi \rightarrow \pi*$ absorption of MYG, and induced only a slight red-shift of about 0.02 eV. Although dispersion effects such as π-stacking can not be expected to be captured correctly at the density functional level, this finding supports the INDO/S calculations in that His197 influences the chromophore excitation mainly through electrostatic interactions. These interactions are captured at the force field level. Furthermore, as enlarging the QM subsystem did not significantly change the results, this control also demonstrates that the size of the original QM subsystem was sufficient for calculating the TDDFT spectra.

To test how the calculated absorption wavelengths depend on the basis set, we repeated the calculations on representative frames from the

MD ensembles using the more complete 6-31+G* and 6-311G* basis sets. The results of these calculations are given in Table 4.1. The absorption energies hardly changed upon increasing the basis set size from 6-31G* to 6-31+G* or 6-311G*.

4.4.2 Protonation States from Continuum Electrostatics Calculations

As an independent check of the protonation state assignment, we computed the protonation probabilities of all titratable groups in the asFP595 protein by Poisson-Boltzmann electrostatics (PBE). These calculations also predict the zwitterion to be the predominant state for the *trans* chromophore in wt asFP595 under physiological conditions, thereby confirming the assignment based on the UV/Vis spectra. The populations of the states Z, N, and A are about 96 %, 3 %, and 1 %, respectively.

To study the effect of *trans*-*cis* isomerization on the protonation pattern in the chromophore binding cavity, we repeated the PBE calculations for both the *trans* and *cis* crystal structures of the mutant. In the mutant, the *trans* isomer is also zwitterionic (population about 99 %), which provides further support for the wt results. For the *cis* conformer, by contrast, the calculations suggest that not the zwitterion, but rather the neutral and the anionic chromophore (states N and A, respectively) are equally populated under physiological conditions. However, in the crystal of the mutant, the cis chromophore is populated to only a minor degree. The weak electron density obtained for the *cis* structure allowed to assign alternative conformations, as compared to the *trans* structure, only for the chromophore and a few additional residues. Therefore, to allow the protein matrix to relax further in response to *trans*-*cis* isomerization, we have repeated the PBE calculations on a *cis* mutant structure that has been energy minimized with positional restraints on all heavy protein atoms (force constant $1000\,\mathrm{kJ\,mol^{-1}nm^{-2}}$). In this slightly relaxed environment, the *cis* chromophore adopts the neutral state with a population of 92 %, while the anionic and zwitterionic states are populated to only 6 % and 2 %, respectively. Thus, *trans*-*cis* isomerization is accompanied by a change in protonation pattern from zwitterionic to neutral, although the anionic *cis* state might be accessible under physiological conditions as well.

Minimization of the *trans* structure prior to the PBE calculations had no effect on the preferred protonation state of the chromophore.

However, His197 was predicted to be doubly protonated in the minimized structure, whereas the calculations from the x-ray structure predicted it to be singly protonated. This suggests that the actual protonation state of His197 depends critically on the local environment in the chromophore pocket. We therefore cannot unambiguously assign the protonation state of this residue. Most probably, both singly and doubly protonated states are populated under physiological conditions. A cationic His197 could stabilize a partially negatively charged chromophore phenolate moiety and might therefore be favored in Z and A states.

The different populations of the protonation states in the *trans* and *cis* conformations are due to the higher acidity of the zwitterionic imidazolinone NH proton in the *cis* isomer compared to that in the *trans* isomer. Our density functional calculations yielded pK values of 4.7 and 9.1 for the *cis* and *trans* chromophores in aqueous solution, respectively. This shift of about 4.4 pK units is due to steric repulsion between the NH hydrogen and one of the phenolate *ortho* protons because these are forced into close proximity in the *cis* isomer.

Our continuum electrostatics estimates of protonation probabilities for the chromophore in the protein matrix depend critically on the accuracy of the reference pK_A-value in aqueous solution. Although it is in principle possible to calculate these values accurate to within half a pK-unit [237], such computations require Complete Basis Set (CBS) methods which are prohibitively expensive for large molecules, such as the asFP595 chromophore. Instead, we used the less accurate density functional calculations described in Section 4.3.3. Assuming similar and systematic errors for both the *cis* and *trans* conformation, we consider the observed differences for the two isomers sufficiently accurate.

In summary, the PBE results corroborate that the chromophore *trans-cis* photoisomerization upon kindling of asFP595 is accompanied by proton transfer events. After or during isomerization, a proton is transferred from the imidazolinone ring of the zwitterion to Glu215, thereby yielding the anionic chromophore. Since the calculations revealed the neutral chromophore (state N) as the most stable species in the *cis* conformation, the anion is probably a metastable intermediate that is readily protonated at the phenolate oxygen to give the neutral *cis* chromophore. Combining the calculated optical absorption spectra and the results of the Poisson-Boltzmann electrostatics allows to proceed further and to address the mechanism of asFP595 kindling in detail. Subsequently, we will focus on the proton wires mediating the interconversion of the chromophore protonation states.

4.4.3 Proton Wires

Protonation of the phenolate oxygen of MYG requires a proton donor in close proximity. Figure 4.1 shows that, intuitively, His197 could provide the proton, but it seems to be too far away for a direct proton transfer. To address this issue, we performed a 20 ns force field MD simulation of the anionic *cis* state and identified relevant hydrogen bonds (data not shown). Indeed, His197 did not form a hydrogen bond to the chromophore in the course of the simulation due to their coplanar arrangement. Instead, a stable hydrogen bond was established between His197 and Glu145 (Figure 4.4), and the glutamic acid might therefore serve as a proton shuttle. However, proton shuttling would additionally require a chain of two or three water molecules to be established between Glu145 and the phenolate oxygen O12, which was not observed in either the MD simulation or in the crystal structure. Figure 4.4 suggests an alternative protonation pathway.

Close examination of the simulations revealed that protons in the chromophore pocket could exchange with protons in the bulk through two distinct proton wires connecting the chromophore cavity to the exterior solution. These wires involve several protonatable side chains and one buried crystallographic water molecule. The proton entrance pathway involves Ser158, which is directly hydrogen bonded to the phenolate oxygen of the chromophore. In the simulations, this residue is connected to the bulk solution through a short chain of water molecules (W3 and W4, Figure 4.4), thereby enabling proton uptake by the chromophore from the solvent. In the crystal structure, the water molecule W233 (Figure 4.1a) is also hydrogen bonded to the phenolate oxygen and could therefore also be involved. The putative proton-release pathway of the His197 proton starts at Glu145, and proceeds through Lys67, Glu195 and Tyr72 towards Asp78. The crystal water molecule W241 temporarily bridges Tyr72 and Asp78 in the simulation, as is shown in Figure 4.4. Located on the surface of the protein, Asp78 finally releases the proton to the exterior solution. The hydrogen bonded network remained intact during the simulations, such that the donor and acceptor distances were always optimal for proton transfer. These results suggest that the proton exchange occurs by rapid one-dimensional diffusion of protons along the two wires.

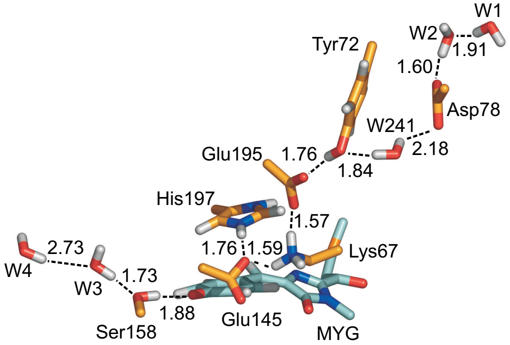


Figure 4.4: Snapshot from an MD simulation showing the proposed proton-entry and -release wires in asFP595. Color code: Carbon atoms of the chromophore (MYG) in cyan and of the amino acid sidechains involved in the proton wires in orange, oxygen atoms in red, nitrogen atoms in blue. Aliphatic hydrogen atoms are not shown. The distances between the proton-donor and -acceptor atoms are given in Å.

4.5 Conclusions

Understanding the kindling mechanism of asFP595 at an atomic level represents a major challenge but is essential for a targeted improvement of this fluoroprotein towards an efficient use as a fluorescence marker in optical microscopy. Structural studies of asFP595 mutants revealed that kindling involves a *trans-cis* isomerization of the covalently bound chromophore, but did not address the crucial role of protons in the kindling process. Here, by calculating optical absorption spectra in combination with Poisson-Boltzmann electrostatics, we elucidated the crucial role of proton transfer processes in the photoswitching mechanism.

Similarly to GFP, the asFP595 chromophore is neutral in the *cis* ground state conformation. Both in asFP595 and GFP, proton wires connect the chromophore to the exterior solution, thereby mediating protonation state changes. These similarities suggest that proton transfers might be essential for the mechanistic understanding also of other fluorescent proteins. Indeed, it was shown recently that the mechanisms of the GFP-like proteins Dronpa [199], DsRed [246], and EosFP [247] involve proton translocations as well. However, for all three cases the molecular mechanisms and proton pathways remain to be elucidated at the atomic level.

The conformational flexibility of the asFP595 chromophore is, in contrast to that of GFP, sufficiently high to allow *trans-cis* isomerization which triggers a proton transfer cascade, preventing the immediate reisomerization to the initial dark state induced by a second photon. Other photoactive proteins have also evolved mechanisms to avoid the immediate photochemical back-reaction. As in asFP595, such mechanisms require that the initial photoproduct is not the most stable ground state minimum, but rather an intermediate in the overall process. Photoisomerization leading to changes in the protonation probabilities is also known as the critical step in the signal transduction mechanism of bacteriorhodopsin [144, 146, 248] and photoactive yellow protein [40, 63, 249]. The quite different structure of these three examples suggests that evolution has exploited this idea as a general principle.

Chapter 5

Photoisomerization of asFP595

In the previous chapter, we have studied the protonation state changes that accompany the photoswitching of asFP595. We found that the positions of the protons in the chromophore cavity crucially determine the spectroscopic properties of asFP595 and elucidated the proton translocation pathways in the protein matrix. In this chapter, we will focus on the excited state dynamics of the *trans*-*cis* photoisomerization process of the chromophore within the protein matrix.

Previously, in close collaboration with Frauke Gräter, we studied the *trans*-to-*cis* isomerization of the asFP595 chromophore by force field MD simulations in which a simplified potential was used to model the excited state potential energy landscape [205]. This simplified excited state potential was based on atomic partial charges derived from configuration interaction singles (CIS) calculations in the gas phase, which was one of my contributions to this previous work. Taken together, the force field simulations revealed that the isomerization follows a space-saving "hula-twist" mechanism, in which the ring-bridging CH group preferably rotates towards His197.

The aim of this chapter is to obtain a even more detailed mechanistic picture of the photoswitching mechanism of asFP595 at the atomic level, i.e., to understand the dynamics of both, the activation process (*off*-to-*on* switching) as well as the de-activation process (*on*-to-*off* switching). The simulations presented in this chapter will address the following key questions: How does light absorption induce the isomerization of the chromophore within the protein matrix, and how do the different protonation states affect the internal conversion mechanism? Which is the fluorescent species, and how can the fluorescence quantum yield be increased?

To address the above questions, hybrid quantum-classical (QM/MM) excited state MD simulations were carried out. By identifying key residues in the chromophore cavity, the simulations elucidate how the protein governs the photoreactivity. Based on the simulations, a complete and new mechanism is provided that explains a number of experimental results and that provides predictions rigorously testable by experimental means.

5.1 Methods

To model the dynamics of the photoactivated asFP595 chromophore, we carried out excited state QM/MM [53] molecular dynamics (MD) simulations. Energies and forces of the excited (S_1) and ground states (S_0) were calculated on-the-fly at the CASSCF/3-21G level of theory [131, 132] with a reduced active space of 6 electrons in 6 orbitals. The studied processes start in S_1 and end up in S_0. The transitions (hops) between the two energy surfaces were modeled by surface selection at the conical intersection (CI) seam. For all MD simulations, Gromacs 3.3 [183] with an interface [63] to Gaussian03 [185] was used. Modifications were made to the one-electron integral routines of Gaussian03 to account for the polarization of the CASSCF wavefunctions due to the point charges on the MM atoms.

The reduced active space used in the MD simulations was validated using geometry optimizations of relevant excited state minima and minimum energy crossing points for the isolated chromophores. The full CASSCF active space for the π-system of the asFP595 chromophore would require 18 π electrons in 16 π orbitals, rendering geometry optimizations prohibitively expensive. To make the respective calculations feasible, the number of excitations in the wavefunction was restricted by employing the RASSCF method [250, 251, 252]. Martial Boggio-Pasqua has characterized minima and conical intersections at the RASSCF(18,7+4+5)[2,2]/6-31G* level of theory. The final 6 electron, 6 orbital active space used in the CASSCF QM/MM MD simulations was selected from the RASSCF calculations such as to enable the simultaneous description of the electronic ground and first excited states.

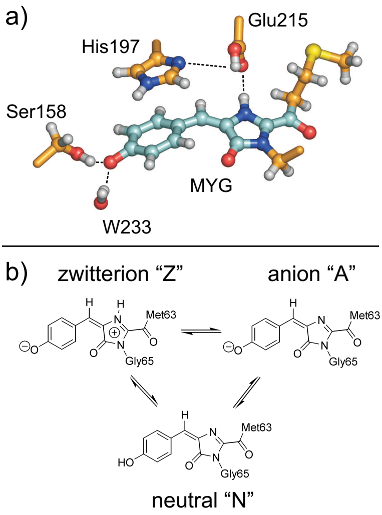


Figure 5.1: **a)** Chromophore (MYG) in the *trans* conformation and adjacent side chains of some amino acids in the chromophore cavity. Glu215, Ser158, and the crystallographic water molecule W233 are hydrogen-bonded to MYG, and His197 is π-stacked to the MYG phenoxy-moiety. The carbon skeleton of the QM subsystem used in the QM/MM simulations is shown in cyan, and the MM carbon atoms are shown in orange. **b)** Schematic drawings of the different MYG protonation states considered in this work.

All MD simulations were based on the crystal structure of a mutant of asFP595. The mutant has similar photochromic properties as the wild-type, but high-resolution crystal structures are available for both the *trans* and the *cis* conformations (see chapter 4). For both *trans* and *cis* conformations, MD simulations were initiated for three different chromophore protonation states; anionic (A), neutral (N), and zwitterionic (Z). The simulations were performed in a rectangular periodic box of about 730 nm^3. Each system contained about 21,500 TIP4P water molecules, including 340 crystallographic water molecules. After assigning the protonation pattern of the chromophore pocket, all other polar, aromatic, and aliphatic hydrogen atoms were added to the protein with the HB2MAK [224] routine of WHATIF [225]. To each of the systems, sodium and chloride ions were added at physiological concentration to compensate for the net positive charge of the protein. The actual number of ions depended on the total charge of the protein, which differed for the chosen protonation patterns of the chromophore cavity. The final systems contained about 90,000 atoms.

Prior to the MD simulations, the systems were energy minimized (1000 steps steepest descent). Subsequently, force field based MD simulations were carried out. First, 500 ps MD simulations with harmonic position restraints on all protein heavy atoms (force constant $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$) were carried out to equilibrate the solvent and the ions. Then, 500 ps free MD were run at 300 K. All simulations were carried out using the OPLS all-atom force field [108]. Parameters for the chromophore are described in chapter 4.

The simulations were run at constant temperature and pressure by coupling to an external heat bath ($\tau_\mathrm{T} = 0.1\,\mathrm{ps}$, $\tau_\mathrm{p} = 1\,\mathrm{ps}$) [187]. In the force field simulations, LINCS [226] was used to constrain bond lengths, thus allowing a time step of 2 fs. SETTLE [188] was applied to constrain the internal degrees of freedom of the water molecules. A twin-range cut-off was used for the Lennard-Jones interactions. Interactions within 1.0 nm were updated every step, whereas interactions between 1.0 nm and 1.6 nm were updated every ten steps. Coulomb interactions within 1.0 nm were computed at each step as well. Beyond this cut-off, the particle-mesh Ewald (PME) method [227] with a grid spacing of 0.12 nm was used.

The QM subsystem in the excited state QM/MM simulations consisted of the chromophore, and the rest of the system was modeled with the OPLS force field (Figure 5.1). The $N - C_\alpha$ bond of Gly65 and the $C_\alpha - C_\beta$ bond of Met63, respectively, were replaced by a constraint, and the QM part was capped with hydrogen link atoms [161]. The forces on the link atoms were distributed over the two heavy atoms at the boundary according to the lever rule. The QM system experienced the Coulomb field of all MM atoms within a sphere of 1.6 nm, and Lennard-Jones interactions between QM and MM atoms were added. In the QM/MM simulations, a time step of 1 fs was used, and no constraints were applied in the QM subsystem. Prior to the excited state simulations, the systems were simulated in the ground state, first for 1 ps at the RHF/3-21G//OPLS level of theory and then for an additional 2.5 ps at the CASSCF(6,6)/3-21G//OPLS level. From the latter trajectory, frames at equal time intervals ($\Delta t = 0.5\,\mathrm{ps}$) were used as starting configurations for the excited state MD simulations.

To accelerate the escape from the S_1 minima, additional excited state conformational flooding simulations (see chapters 2 and 3) were carried out for the N_{cis}, Z_{cis}, and Z_{trans} systems. A gaussian-shaped flooding potential V_fl was constructed from a principal component analysis (PCA) [176, 177, 178] of free S_1 simulations. For all systems, the covariance matrix of the motion of all QM atoms except the exocyclic carbonyl group of the chromophore and the hydroxyphenyl-OH proton (for N_{cis}) was computed from two independent 10 ps excited state trajectories. All internal degrees of freedom of the chromophore were affected by the flooding potential to ensure that the escape from the initial minimum was accelerated in an unbiased manner. Unless stated differently, adaptive flooding with target destabilization free energies of 300 kJ/mol (Z_{trans}) or 100 kJ/mol (N_{cis}) and time constants of $\tau = 0.1\,\text{ps}$ was applied. In all flooding simulations, V_{fl} was switched off after the conical intersection seam was encountered to allow for an unperturbed relaxation on the ground state potential energy surface.

5.2 Results and Discussion

Our results reveal that the excited state behavior of asFP595 is determined by the protonation pattern of the chromophore and some amino acids in the surrounding protein matrix rather than by the chromophore conformation (*trans* or *cis*). The latter, however, modulates the excited state properties by changing the hydrogen-bonded network in the chromophore cavity.

For both conformers, we identified three possible protonation species, explaining the complex photochemical behavior of asFP595. First, the neutral species N_{trans} and N_{cis} undergo reversible *trans*-*cis* photoisomerization and thus account for the photoswitching between the dark *off* and fluorescent *on* states. Second, anionic chromophores A_{trans} and A_{cis} lead to the observed ultra-fast radiationless deactivation. Third, fluorescence emission can in principle originate from both zwitterions Z_{trans} and Z_{cis}. The protonation species are interchangeable *via* proton transfers. In the following we will describe the excited state behavior of all three protonation species.

5.2.1 *Trans-Cis* Isomerization of the Neutral Chromophore

Five excited state simulations were initiated from the ground state trajectory of the *trans* neutral chromophore N_{trans}. In these simulations, excited state (S_1) lifetimes between 0.224 ps and 0.718 ps were observed (Table 5.1). A simple exponential fit to the observed lifetimes yielded a decay time of $\tau = 0.37 \pm 0.01\,\text{ps}$. Recent femtosecond time-resolved pump/probe experiments by Schüttrigkeit and coworkers have yielded excited state decay time constants of 0.32 ps (78%), 2.6 ps (19%), and 12.1 ps (3%) as well as a fluorescence lifetime of 2.2 ns for asFP595 [9]. However, although the simulated decay times seem to be in good agreement to the experimental results, we believe that the results cannot be

starting in N_{trans}		
run	S_1 lifetime (ps)	final conformation
a	0.516	trans
b	0.475	cis
c	0.309	trans
d	0.224	trans
e	0.718	trans

Table 5.1: Excited state lifetimes and final conformations from the MD simulations initiated in the neutral *trans* chromophore conformation.

compared. In chapter 4 it was demonstrated that the N_{trans} species is hardly populated in asFP595 and therefore cannot contribute to the observed excited state decay. Instead, the species that is predominantly responsible for the ultra-fast radiationless decay observed in the experiments is the anionic *trans* chromophore A_{trans}, as shown in detail below. *Trans*-to-*cis* photoisomerization of the chromophore was observed in one of the simulations (run b, Table 5.1). Figure 5.2 shows a schematic representation of the S_0 (green) and S_1 (red) potential energy surfaces of the neutral chromophore, along with a photosiomerization MD trajectory (yellow dashed line). The multi-dimensional surfaces are projected onto the isomerization coordinate and a skeletal deformation coordinate of the imidazolinone moiety (see below). The dynamics can be separated into three distinct phases: (i) evolution on the electronic ground state S_0, (ii) excitation and evolution on the excited state S_1, and (iii) decay back to S_0 at the surface crossing seam followed by subsequent relaxation on the ground state surface. The position of the surface crossing seam controls the passage of the trajectory from S_1 to S_0. The seam is accessed from a global twisted minimum on S_1, which is separated by a small S_1 barrier from a local planar minimum near the Franck-Condon region. Figure 5.3a shows the snapshot from the isomerization trajectory shortly before the surface crossing seam was encountered. After photon absorption (blue arrow in Figures 5.2 and 5.3b), the chromophore spontaneously rotated around torsion A (imidazolinone-twist), and the ring-bridging CH group pointed downwards (away from His197) by almost 90°. The time-evolution of the S_0 and S_1 potential energies and of the two ring-bridging torsion angles during *trans*-*cis* photoisomerization

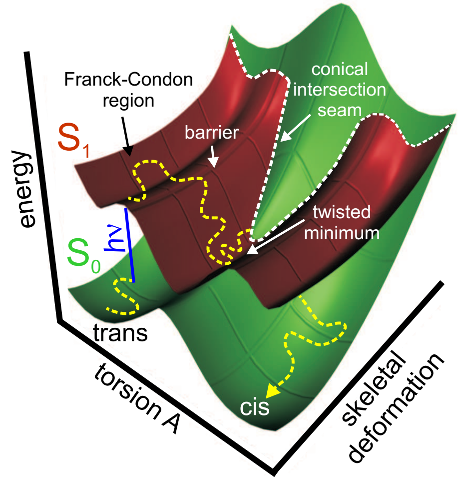


Figure 5.2: Schematic representation of the potential energy surfaces of the excited (S_1, red) and ground (S_0, green) states of the neutral chromophore along the *trans-cis* isomerization coordinate (torsion A) and a skeletal deformation coordinate of the imidazolinone moiety. The dashed yellow line represents the path sampled in a QM/MM photoisomerization trajectory. Radiationless decay occurs at an extended S_1/S_0 conical intersection seam (dashed white line).

are shown in Figures 5.3b,c. After excitation to S_1, the chromophore rapidly relaxed from the Franck-Condon region into a nearby planar S_1 minimum, as is evident from the decreasing S_1 energy in panel b (red curve). The system stayed in this planar minimum for about 0.2 ps, before the global twisted S_1 minimum was adopted through rotation around torsion A (Figures 5.2 and 5.3b,c). The system oscillated around this minimum until the conical intersection seam was encountered. After the surface hop to S_0, rotation around torsion B (hydroxyphenyl-twist) followed with a short delay of about 0.5 ps. Previously, an ideal "hula-twist" isomerization mechanism was proposed [205], which would involve a *simultaneous* rotation around *both* torsion angles A and B. Although rotation around both torsions was observed in our simulations, the torsional twisting occurred consecutively rather than simultaneously. Thus, the isomerization proceeded *via* the twisted conformer with perpendicular imidazolinone and hydroxyphenyl moieties shown in Figure 5.3a.

During the initial equilibration of the N_{trans} species, the hydrogen bonding network obtained from the x-ray crystal structures of the anionic and zwitterionic chromophores (see above) was altered to accommodate the non-native neutral chromophore. First, a stable hydrogen bond was formed between the hydroxyphenyl OH group of MYG and Glu145 (Figure 5.3a). Second, the hydrogen bonds between the imidazolinone nitrogen and Glu215 as well as between His197 and Glu215 broke. Interestingly, these two hydrogen bonds were transiently re-established during the end of the second isomerization phase (blue and orange curves in Figures 5.3d and e, respectively) in which torsion B followed torsion A. In the twisted intermediate structure, the imidazolinone nitrogen atom was sterically more exposed as compared to the planar conformation, which facilitated the formation of the hydrogen bonds. During an extended 10 ns force field simulation, the His197–Glu215 and MYG–Glu215 hydrogen bonds repeatedly broke and re-formed at a timescale of several hundred picoseconds (data not shown). Figures 5.3d,e furthermore show that during the isomerization, the hydrogen bonding network in the chromophore cavity was stable as none of the hydrogen bonds established at the instant of photoexcitation ruptured. A similar stability was found in all simulations, irrespective of the chromophore conformation or protonation state.

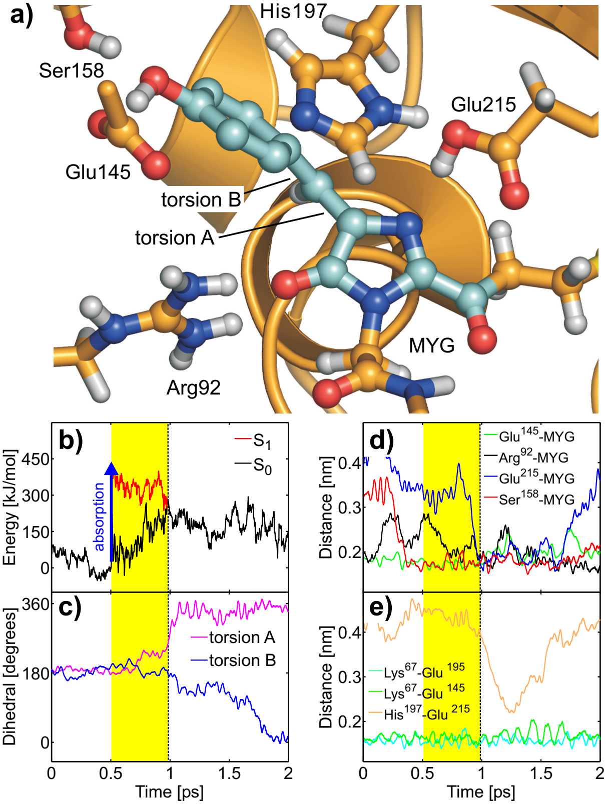


Figure 5.3: *Trans-cis* isomerization of the neutral chromophore. **a)** Chromophore (MYG) conical intersection geometry adopted during the MD simulation. MYG forms hydrogen bonds to Arg92, Glu145, Ser158, and Glu215. Color code as in Figure 5.1 **b)** Ground (S_0, black) and excited (S_1, red) potential energy traces along the QM/MM molecular dynamics trajectory. Photon absorption (green arrow) excites the chromophore into S_1 (yellow area) until it decays back to S_0 at the conical intersection seam (dashed line). **c)** Time-evolution of the ring-bridging torsion angles A (magenta) and B (blue). **d,e)** Hydrogen bonding network in the chromophore cavity during *trans-cis* isomerization. The MYG–Arg92 (black), MYG–Glu145 (magenta), MYG–Ser158 (red), Lys67–Glu195 (cyan, residues not shown in **a**), and Lys67–Glu145 (green) hydrogen bonds were stable during isomerization. Additional hydrogen bonds between MYG and Glu215 (blue) as well as between His197 and Glu215 (orange) were transiently formed.

starting in N_{cis}		
run	S_1 lifetime (ps)	final conformation
a	0.374	trans
b	3.561	cis
c*	1.573	trans
d*	0.867	cis
e*	1.206	trans

Table 5.2: Excited state lifetimes and final conformations from the MD simulations initiated in the neutral *cis* chromophore conformation. *In runs c,d, and e, the escape from the S_1 minimum was accelerated by conformational flooding.

For the neutral *cis* chromophore N_{cis}, five excited state simulations were initiated from the ground state trajectory (Table 5.2). Only two of these trajectories returned to the ground state within 10 ps (Table 5.2, runs a and b), which was considered the maximum affordable trajectory length in terms of computation time (about two CPU-months). In one of these two simulations, a spontaneous *cis*-to-*trans* photoisomerization was observed (run a). As expected, the isomerization pathway was similar to the reverse *trans*-to-*cis* pathway in that the conical intersection seam was accessed *via* rotation around torsion A, followed by a slightly delayed rotation around torsion B in S_0. However, in contrast to the activation pathway, the ring-bridging CH group rotated upwards (i.e., towards His197). Thus, despite the anisotropic protein surrounding, both rotational orientations of the chromophore CH bridge are feasible. In the second simulation, the CI seam was also encountered after rotation around torsion A, but the chromophore returned to the initial *cis* conformation.

In the other three trajectories, the chromophore remained trapped in a planar S_1 minimum conformation near the Franck-Condon region throughout the simulation (not shown in Table 5.2). The starting structures of these trajectories were used for three additional simulations in which the escape from the planar S_1 minimum was accelerated by means of conformational flooding (Table 5.2, runs c–e). In these accelerated simulations, the flooding potential successfully induced the escape from the local S_1 minimum, and the surface crossing seam was encountered in all cases. Isomerization was observed in two of these simulations. In total, *cis*-to-*trans* photoisomerization was seen in three out of five simulations initiated in the N_{cis} state. Although the number of trajectories is statistically low, our simulations suggest that the probability for *cis*-to-*trans* isomerization is higher than for the reverse *trans*-to-*cis* process (see above).

S_1/S_0 Conical Intersection Topology and S_1 Minima

To further characterize the potential energy surfaces underlying the photochemical conversion processes of the neutral chromophore (Figure 5.2), three S_1 minima and a minimum energy S_1/S_0 conical intersection (MECI) were optimized in the gas phase. A local planar S_1 minimum for the *trans* isomer lies 76 kJ/mol below the Franck-Condon (FC) region. The global S_1 minimum is twisted around torsion A by 85° and lies -131 kJ/mol relative to the FC region. The nearby MECI is twisted around torsion A by 81°. The MECI is energetically lower than the Franck-Condon region by 62 kJ/mol, and the CI seam is therefore readily accessible. Twisting around torsion B instead of torsion A also leads to a local minimum on S_1, whose energy is 28 kJ/mol below the FC geometry.

The MD simulations reflect this surface topology. Immediately after the excitation, the system relaxed from the FC region to the global S_1 minimum by rotation around torsion A (Figure 5.3c). The system oscillated around this minimum until the conical intersection seam was encountered and a surface hop back to S_0 took place. The gradients on S_0 and S_1 are almost parallel at the CI, which indicates that the CI is sloped. The gradient difference vector and the derivative coupling vector that span the branching space largely correspond to skeletal deformations of the imidazolinone moiety. Thus, as shown in Figure 5.2, the rotation coordinate around torsion A or B is parallel to the seam and does not lift the S_1/S_0 degeneracy. The seam is accessible anywhere along the torsional rotation coordinates, and therefore such torsional rotation is in principle not essential for the radiationless decay. The extended surface crossing seam parallel to the isomerization coordinate accounts for the low isomerization quantum yield seen in our simulations. In the majority of our MD simulations, the seam was encountered rather "early" along the torsional rotation coordinate (Figure 5.2), and the system thus returned to the ground state before overcoming the S_0 barrier maximum. In these cases, relaxation on S_0 after the surface hop led back to the starting conformation.

Role of the Protein Environment

To elucidate the role of the protein environment in the photoisomerization process of the chromophore, we have re-calculated the S_1 and S_0 energies along two excited state trajectories (run b, Table 5.1 and run a, Table 5.2) in the gas phase. In these simulations, the chromophore followed the same trajectory as before, but did not interact with the rest of the system (protein and solvent).

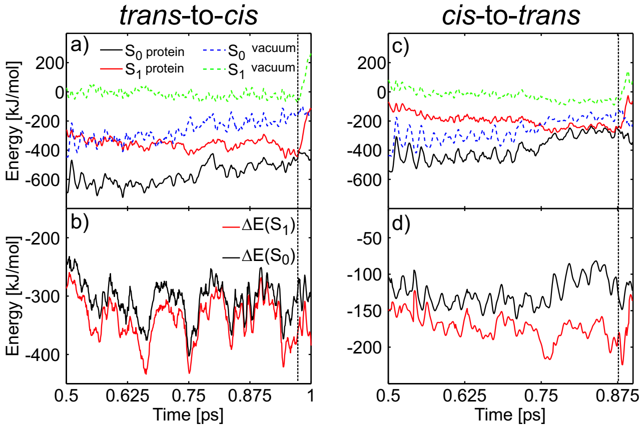


Figure 5.4: Influence of the protein environment on the photoisomerization process of the neutral asFP595 chromophore. **a,c)** Ground and excited state energies along *trans*-to-*cis* (a) and *cis*-to-*trans* (c) isomerization trajectories (run b, Table 5.1 and run a, Table 5.2). The protein environment stabilizes S_0 and S_1 (black and red lines, respectively) relative to the gas phase (dashed blue and green lines, respectively). **b,d)** Energy difference between the protein and the gas phase. $\Delta E\,(S_0) = E\,(S_0, \text{protein}) - E\,(S_0, \text{gas phase})$ is plotted in black, $\Delta E\,(S_1) = E\,(S_1, \text{protein}) - E\,(S_1, \text{gas phase})$ in red. The protein environment energetically stabilizes S_1 more strongly than S_0. The vertical dashed black line represents the surface crossing. The energy offset in (a) and (c) is 1.9699×10^6 kJ/mol.

Figures 5.4a,c show the obtained energy traces. In the protein, both S_1 and S_0 are stabilized with respect to the gas phase. For the *trans*-to-*cis* isomerization process, the protein stabilized the energies of the S_1 and S_0 states on average by $-339\,\text{kJ/mol}$ and $-307\,\text{kJ/mol}$, respectively. For the *cis*-to-*trans* process, the average stabilization energies were $-173\,\mathrm{kJ/mol}$ and $-126\,\mathrm{kJ/mol}$, respectively. Thus, the protein (and solvent) environment stabilized S_1 stronger than S_0 by about 30–50 kJ/mol. The energy differences between the protein and the gas phase, $\Delta E = E\,(\text{protein}) - E\,(\text{gas phase})$ are plotted in Figures 5.4b,d. The S_1 stabilization was rather strong at the surface crossing seam (dashed lines in Figure 5.4). We found S_1 to be stabilized stronger than S_0 by 78 kJ/mol and 93 kJ/mol at the conical intersection in both MD simulations. In summary, the protein environment energetically stabilizes S_1 stronger that S_0, thereby facilitating fast radiationless decay.

5.2.2 Ultra-fast Radiationless Deactivation of the Anionic Chromophore

In total, 20 simulations of the anionic chromophore protonation state were carried out, 10 of which were initiated in the *trans* conformation and the other 10 were initiated in the *cis* conformation. The individual excited state lifetimes in Table 5.4 show that ultra-fast radiationless deactivation occurred in all 20 trajectories. However, *trans-cis* photoisomerization was never observed. A simple exponential fit to the S_1 lifetimes of the *trans* anion yielded a decay time of $\tau = 0.47 \pm 0.02\,\mathrm{ps}$. Since A_{trans} is one of the two dominant protonation species in the *off* state, besides Z_{trans}, we expect it to significantly contribute to the experimentally observed decay. The measured decay time of 0.32 ps [9] agrees well with the decay time from the simulations. For A_{cis} an excited state decay time of $\tau = 2.29 \pm 0.04\,\mathrm{ps}$ was obtained, which is about five times longer as compared to the decay time of A_{trans}.

Figure 5.5a shows the conical intersection geometry adopted during a typical trajectory. In contrast to the neutral chromophore, the CI seam was accessed through a phenoxy-twist (rotation around torsion B, see Figure 5.5c), and the CH bridge remained in the imidazolinone plane. Shortly after the excitation, rotation around torsion B drove the system towards the surface crossing seam (Figures 5.5b,c). Back on S_0, the system returned to the initial configuration. The hydrogen bonding network in the chromophore cavity was very similar to the network observed in the x-ray crystal structures and remained stable during the excited state MD simulations.

Since rotation around torsion B does not lead to *trans-cis* isomerization and rotation around torsion A did not occur, the quantum yield for isomerization was zero in our simulations. However, due to the limited

run	S_1 lifetime (ps)	initial conformation	final conformation
a	0.363	trans	trans
b	0.483	trans	trans
c	1.190	trans	trans
d	0.461	trans	trans
e	0.508	trans	trans
f	3.141	cis	cis
g	1.997	cis	cis
h	0.731	cis	cis
i	2.811	cis	cis
j	3.451	cis	cis
k	0.581	trans	trans
l	0.565	trans	trans
m	0.493	trans	trans
n	0.837	trans	trans
o	0.300	trans	trans
p	3.743	cis	cis
q	4.900	cis	cis
r	1.176	cis	cis
s	2.965	cis	cis
t	1.750	cis	cis

Table 5.4: Excited state lifetimes and conformations from the MD simulations of the anionic chromophores A_{trans} and A_{cis}. In runs a–j, His197 was modeled cationic, whereas it was modeled as neutral (protonated at N_δ) in runs k–t.

number of trajectories (20), we cannot exclude the *trans-cis* photoisomerization of the anionic species. Our results agree with recent MRPT2 computations of Olsen and coworkers on an anionic DsRed-like model chromophore in the gas phase, who have shown that the imidazolinone-twisted S_0/S_1 CI (i.e., twisted around torsion A), which leads to *cis-trans* isomerization, is more than 150 kJ/mol higher in energy as compared to the phenoxy-twisted CI [253].

The difference between the S_1 lifetimes of the *cis* and *trans* conformers can be attributed to the steric constraints imposed by the protein matrix. In the *trans* conformation the phenoxy-ring deviates from planarity by as much as 20°, whereas the *cis* chromophore is essentially planar (see chapter 4). For the trans conformer to reach the CI, only a slight additional twisting was required to access the surface crossing seam. Thus, the pre-twisting of the phenoxy-moiety due to the protein matrix facilitated fast internal conversion in our simulations of the *trans* conformer.

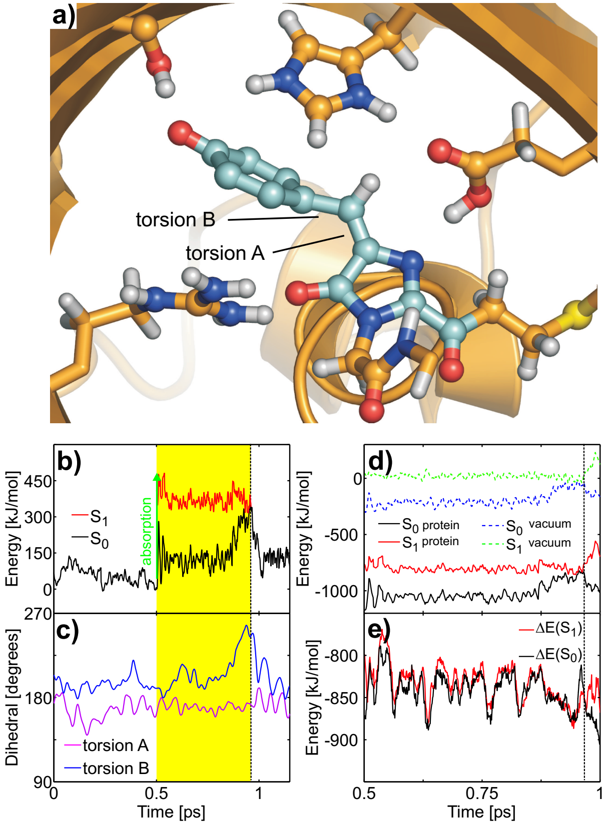


Figure 5.5: Ultra-fast internal conversion mechanism of the *trans* anion. **a)** At the conical intersection snapshot, the chromophore is twisted around torsion B, yet the hydrogen bonded network in the chromophore cavity remains intact. **b)** Ground (S_0, black) and excited (S_1, red) potential energy traces along the QM/MM molecular dynamics trajectory. Photon absorption (green arrow) brings the chromophore into S_1 (yellow area) until it decays back to S_0 at the conical intersection seam (dashed line). **c)** Time-evolution of the torsion angles A (magenta) and B (blue). **d)** S_0 and S_1 energies along a representative excited state trajectory of A_{trans}. The protein environment strongly stabilizes S_0 and S_1 (black and red lines, respectively) relative to the gas phase (dashed blue and green lines, respectively). The energy offset is 1.9686 $\times 10^6$ kJ/mol. **e)** Energy difference ΔE between the protein and the gas phase for S_0 (black) and S_1 (red).

S_1/S_0 Conical Intersection Topology and S_1 Minima

Martial Boggio-Pasqua has optimized the S_1/S_0 MECI, a planar, and two twisted S_1 minima (imidazolinone-twist and phenoxy-twist) for an isolated anionic chromophore. The planar minimum lies 31 kJ/mol below the FC geometry. The global S_1 minimum is twisted around torsion B by 269.1° and lies -82 kJ/mol relative to the FC region. The nearby MECI is twisted about torsion B by 269.4° and is energetically lower than the FC geometry by 61 kJ/mol, explaining the ultra-fast decay seen in our MD simulations. Twisting around torsion A leads to a local S_1 minimum that is 46 kJ/mol below the FC geometry. The CI of the anion is sloped, and the gradient difference vector corresponds to a skeletal deformation of the imidazolinone ring, analogous to the neutral chromophore (see above). In contrast to the neutral chromophore, the derivative coupling vector involves rotation around torsion B. However, the amplitude of this vector is very small. Thus, the two electronic states remain very close in energy along torsion B, allowing the system to decay at various phenoxy-twist angles.

Role of the Protein Environment

To study the influence of the protein matrix on the deactivation process of the anionic chromophore, we have re-evaluated the S_0 and S_1 energies along two representative excited state trajectories (*trans* and *cis*) with all interactions between the QM atoms of the chromophore and the MM surrounding switched off, as done previously for the neutral species (see above). As Figures 5.5d,e show, the protein (and solvent) environment stabilized the chromophore with respect to the gas phase. The S_0 and S_1 states of the *trans* chromophore are strongly stabilized by $-840\,\mathrm{kJ/mol}$ and $-832\,\mathrm{kJ/mol}$, respectively. The S_0 and S_1 states of the *cis* chromophore were stabilized relative to the gas phase by $-407\,\mathrm{kJ/mol}$ and $-433\,\mathrm{kJ/mol}$, respectively, during a representative A_{cis} trajectory (data not shown). Similar to the neutral chromophore the protein environment prefers the *trans* conformation of the anion over *cis*. Before reaching the CI seam, the S_0 and S_1 states of the chromophore were stabilized to the same extent. At the CI, however, the protein environment lowered the energy of the S_1 state more strongly than the energy of the S_0 state by 26 kJ/mol and 20 kJ/mol for A_{trans} and A_{cis}, respectively. This preferential stabilization of S_1 enhanced the ultra-fast radiationless deactivation.

5.2.3 Fluorescence Emission of the Zwitterionic Chromophores

Ten simulations were carried out for the zwitterionic species. Five simulations were started in the Z_{trans} conformation, and the other five simulations were initiated in the Z_{cis} conformation. No decay back to the ground state was observed within a maximum trajectory length of 10 ps, neither for Z_{trans} nor for Z_{cis}. In none of the excited state simulations, the chromophore escaped from a planar S_1 minimum in the vicinity of the Franck-Condon region. This suggests that Z_{trans} and Z_{cis} could be the fluorescent species in asFP595, although the measured fluorescence lifetime of 2.2 ns [9] is still orders of magnitude longer than our maximal trajectory length (10 ps).

For Z_{trans} and for Z_{cis}, we have carried out three additional simulations, respectively, in which we applied the conformational flooding technique to accelerate the escape from the S_1 minimum. Figure 5.6 shows the results obtained for Z_{cis}. Similar results were obtained for Z_{trans} (data not shown). Starting in the planar S_1 *cis* minimum, the flooding potential V_{fl} induced an isomerization in the S_1 state to the *trans* conformation within less than 0.5 ps. During the isomerization the S_0/S_1 energy gap decreased significantly (Figure 5.6b), but no spontaneous surface crossing back to S_0 was observed. Figure 5.6a shows the structure along an isomerization trajectory at which the S_0/S_1 energy gap was minimal. The chromophore adopted a hula-twist conformation with the ring-bridging CH group pointing upwards (towards His197). In the hula-twist mechanism, both ring-bridging torsion angles rotated simultaneously, as shown in Figure 5.6c. After excitation to S_1, the flooding potential V_{fl} (inset of Figure 5.6b) induced the crossing of the S_1 barrier that separates the planar minimum from the hula-twisted conformation. Upon the barrier crossing, the chromophore escaped the influence of the flooding potential and V_{fl} dropped to (almost) zero. We imposed a surface hop at the structure with the minimal S_0/S_1 energy gap and switched off the flooding potential to allow an unperturbed relaxation on the S_0 surface (dashed line in Figure 5.6).

The results thus obtained for the zwitterionic chromophore suggest that a hula-twist CI may be spontaneously accessed if the trajectories were extended to (significantly) longer times. In the next paragraph, we characterize the CI and show that the minimum energy crossing point for the zwitterionic chromophore has a high energy, thus hampering radiationless decay.

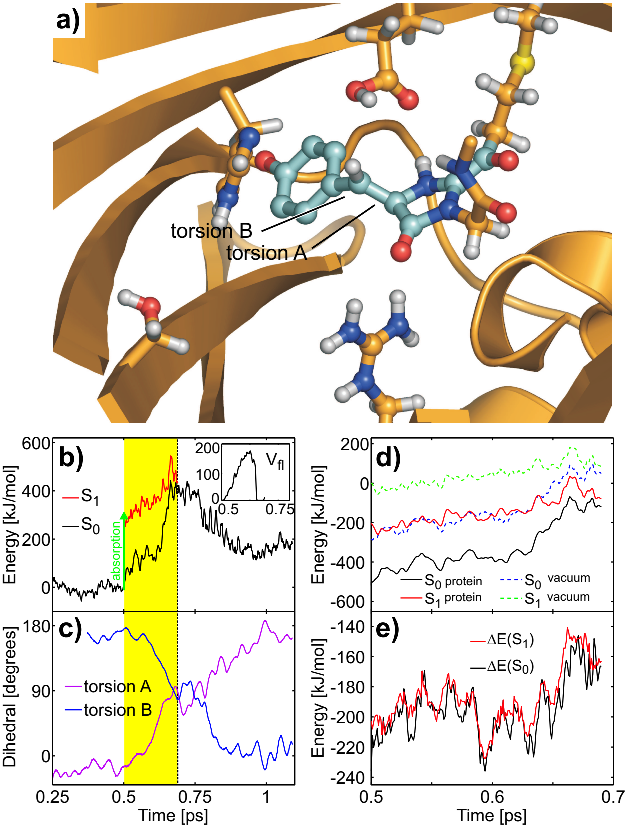


Figure 5.6: Isomerization of the *cis* zwitterion induced by conformational flooding. **a)** Hula-twist structure adopted during isomerization trajectory **b)** Ground (S_0, black) and excited state (S_1, red) potential energy traces along the trajectory. S_0 and S_1 come energetically close, but the surface crossing seam was not encountered. The surface hop was therefore imposed at the structure with the minimum energy gap (dashed black line). The time evolution of the flooding potential V_{fl} is shown in the inset. **c)** Time evolution of the torsion angles A (magenta) and B (blue). **d)** S_0 and S_1 energies (black and red lines, respectively) along the isomerization trajectory. The protein environment stabilizes S_0 and S_1 relative to the gas phase (dashed blue and green lines, respectively). The energy offset is 1.97×10^6 kJ/mol. **e)** Energy difference ΔE between the protein and the gas phase for S_0 (black) and S_1 (red). S_0 is stabilized slightly stronger than S_1 along the whole isomerization pathway.

S_1/S_0 Conical Intersection Topology and Influence of the Protein Environment

A planar S_1 minimum and a hula-twist S_0/S_1 MECI were optimized for the isolated zwitterion. In contrast to the anion and the neutral chromophore, no twisted S_1 minima were found. The gradient difference vector and the derivative coupling vector at the MECI do not involve torsional rotation of either torsion A or torsion B, indicating that the CI seam lies parallel to the isomerization coordinate. The MECI lies 70 kJ/mol above the planar S_1 minimum and 23.4 kJ/mol above the Franck-Condon geometry. Hence, in contrast to the anion and the neutral chromophore, no low-lying CI is present for the zwitterion, demonstrating that radiationless decay in the gas phase cannot occur in an unactivated manner. For the CI seam to become accessible, a significant stabilization of S_1 relative to S_0 by the protein environment would be required. However, as shown in Figures 5.6d,e the protein surrounding does not reduce the S_0/S_1 energy gap anywhere along the isomerization coordinate.

Deactivation of Z_{trans} through Proton Transfer

These results suggest that the zwitterionic chromophore is potentially fluorescent, irrespective of the conformation. However, from the x-ray crystal structures we know that only the *cis* chromophore fluoresces, whereas the *trans* chromophore is dark [205]. A possible explanation for this discrepancy is the presence of an alternative deactivation channel that does not involve isomerization. This deactivation pathway is more easily accessible for Z_{trans} than for Z_{cis}. Only the latter is therefore trapped in S_1 and fluoresces.

The hydrogen bond between the NH group of the imidazolinone ring and Glu215 strongly suggests that the alternative decay involves an excited state proton transfer (ESPT). Such ESPT would quench the fluorescence, because the resulting anion rapidly deactivates, as shown in Section 5.2.2. However, by including only the chromophore into the QM subsystem, we have excluded the possibility of observing such ESPT in our QM/MM simulations. To study the possible effect of the chromophore conformation on the hydrogen bonding network, we have performed extended force-field MD simulations of both Z_{trans} and Z_{cis} and analyzed the relevant hydrogen bonds. As shown in Figure 5.7a, during the simulation of Z_{trans} two stable hydrogen bonds were formed between the protonated OH group of Glu215 and His197 as well as between the

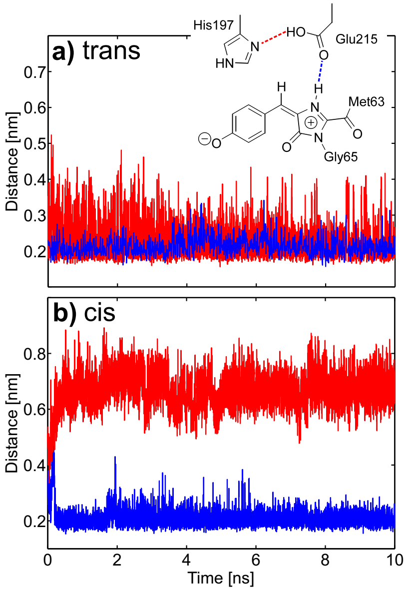


Figure 5.7: Distance between the NH proton of MYG and Glu215 (blue) as well as between OH-group of Glu215 and the N_δ atom of His197 (red) during force-field MD simulations of Z_{trans} (a) and of Z_{cis} (b). The inset in a) shows the corresponding distances.

NH proton of MYG and Glu215. Along these two hydrogen bonds, concerted proton transfer from Z_{trans} to A_{trans} is possible. The OH proton of Glu215 could transfer to the N_δ atom of His197, which is followed by or occurs simultaneous with the transfer of the NH proton of the imidazolinone moiety to Glu215. In the force-field simulation of Z_{cis} the hydrogen MYG–Glu215 bond remained intact, whereas the Glu215–His197 hydrogen bond broke up after about 1 ns (Figure 5.7b). The same results were found in two additional statistically independent MD simulations of Z_{trans} and Z_{cis} (data not shown). Based on these results, we conclude that only the *trans* zwitterion can be converted to the anion through a short proton wire. Therefore, an ultra-fast deactivation channel is available only for the *trans* zwitterion, and not for the fluorescent *cis* zwitterion. Studying the proton hops along the identified pathways in asFP595, both in the ground and the excited state, is intriguing but beyond the scope of the present work.

Having established that fluorescence can only originate from the zwitterionic chromophores, we can now predict the structure of the irreversibly fluorescent state of asFP595. We expect that intense irradiation over a prolonged period of time leads to a decarboxylation of the

Glu215 side chain (Figure 5.8). Such process is also known to occur in GFP [254, 255] and DsRed [246]. A decarboxylated Glu215 can no longer take up the NH proton from the zwitterionic chromophores. The absence of an S_1 deactivation channel leads to fluorescence. The experimental finding that the irreversibly fluorescent state cannot be switched *off* by light (see Introduction) is underlined by our observation that even in the flooding-induced isomerization trajectories, there was no radiationless decay back to S_0.

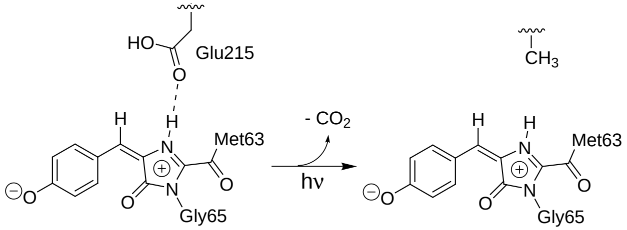


Figure 5.8: Scheme of the proposed decarboxylation of Glu215, which yields an irreversibly fluorescent zwitterion.

5.2.4 Influence of π-stacked His197

Due to its close proximity to the chromophore and its co-planar orientation, we expected the charge state of His197 to influence on the excited state properties of the asFP595 chromophore. The Poisson-Boltzmann electrostatics calculations presented in chapter 4 have shown that slight structural changes in the local environment change the preferred protonation of the His197 imidazole ring between cationic and neutral and that both protonation states are populated at room temperature.

In all our simulations of the anionic and zwitterionic chromophores we considered both states of the His197, i.e., we ran the same number of trajectories with a cationic and with a neutral imidazole ring. In all cases, the His197 protonation state did not influence the decay mechanism. Furthermore, we found that the S_1 lifetimes were hardly affected. For A_{trans}, decay times of $\tau = 0.50 \pm 0.02$ ps and $\tau = 0.45 \pm 0.02$ ps were obtained from the simulations with a cationic and a neutral His197, respectively. For A_{cis}, the respective decay times were $\tau = 2.09 \pm 0.03$ ps and $\tau = 2.48 \pm 0.04$ ps. A closer analysis of the hydrogen bonds in the chromophore cavity revealed the reason for the similar behavior. The protonated N_ϵ of the cationic His197 imidazole ring formed a hydrogen bond to the anionic Glu145 side chain, which reduced the positive charge density due to the cationic His197 above the phenoxy-ring of MYG. In case of the neutral His197, a hydrogen bond between the deprotonated N_ϵ of His197 and the cationic ammonium group of Lys67 was established, leading to the location of the positive Lys67 charge above the chromophore. Thus, in both cases, the electrostatic environment of the chromophore is similar, explaining the similar excited state lifetimes.

5.2.5 Switching Efficiency of asFP595

In the dark resting state, the chromophore adopts the *trans* conformation. To enter the fluorescent *on* state, the chromophore has to isomerize from *trans* to *cis*. Photoisomerization was only observed for the neutral form of the chromophore ($N_{trans} \rightarrow N_{cis}$). However, the N_{trans} state is only marginally populated (see chapter 4), explaining the low quantum yield for switching asPF595 to the *on* state. Moreover, green light is used to switch on asFP595, whereas the absorption maximum of N_{trans} is significantly blue-shifted. The use of blue light, however, would lead to an unfavorable back-reaction due to the absorption of the blue light by N_{cis}, leading to reverse *cis*-to-*trans* isomerization. N_{cis} is the predominant protonation species in the *cis* conformation, thus explaining the high efficiency for the *on*-to-*off* switching. Fluorescence originates from the Z_{cis} species, which like N_{trans} is also hardly populated (see chapter 4). Taken together, the low populations of the involved states give rise to the low overall fluorescence quantum yield of asFP595.

The insight thus obtained from our simulations can be exploited for a targeted improvement of asFP595 for applications as a fluorescence marker in optical microscopy. In particular, to improve the signal-to-noise ratio a higher fluorescence quantum yield is desired. One way to enhance fluorescence is to increase the stability of Z_{cis}, e.g., by introducing additional hydrogen bond donors near the phenoxy-group of the chromophore. Another possibility is to implement an internal proton relay, similar to that in GFP. In GFP, a hydroxyphenyl-bound serine residue, a water molecule, and a glutamic acid form an internal proton wire that enhances the formation of the fluorescent A_{cis} species from the neutral chromophore *via* excited state proton transfer. GFP has a significantly higher fluorescence quantum yield as compared to asFP595 [149, 148, 147, 209, 210]. Although the fluorescent species in

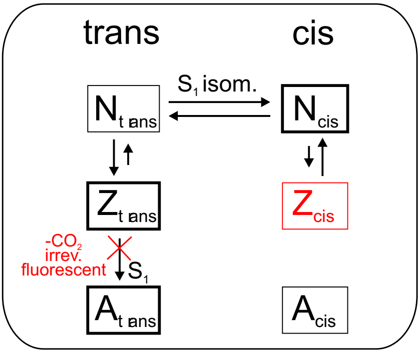


Figure 5.9: Scheme of the reversible photoswitching mechanism of asFP595 proposed in this work. The fluorescent state Z_{cis} is highlighted in red. The major protonation species are indicated by the bold boxes.

asFP595 and GFP are different, the similarity between the chromophores suggests that implementing a similar internal proton relay in asFP595 might increase the fluorescence quantum yield. However, due to the competition between different reaction channels in asFP595, shifting the relative populations of the protonation species will also affect the photoswitchability. For example, increasing the population of the fluorescent species at the cost of the neutral species will decrease the back-isomerization efficiency. Thus, a compromise has to be found between increasing the fluorescence quantum yield on the one hand while maintaining the photoswitchability of asFP595 on the other hand.

5.3 Conclusions

Understanding the excited state dynamics of ultra-fast photoactivated processes in biomolecular systems such as the reversible photoswitching of the fluorescent protein asFP595 represents a major challenge, but is essential to unveil the underlying molecular mechanisms. The present work demonstrates that by using an *ab initio* QM/MM excited state molecular dynamics strategy together with explicit surface hopping, it is not only possible to explain experimental quantities such as quantum yields and excited state lifetimes, but also to make predictions that are rigorously testable by experimental means, such as the nature of the irreversibly fluorescent state or possible improvement to the fluorescence quantum yield.

Figure 5.9 summarizes the proposed photoswitching mechanism. The proton distribution at the active site of asFP595 governs the photochemical conversion pathways of the chromophore in the protein matrix.

Changes in the protonation state of the chromophore and some proximal amino acids lead to different photochemical states, which are all involved in the photoswitching process. These photochemical states are (i) the neutral chromophore species N_{trans} and N_{cis}, which can undergo *trans-cis* photoisomerization, (ii) the anionic chromophores A_{trans} and A_{cis}, which rapidly undergo radiationless decay after excitation, and (iii) the potentially fluorescent zwitterions Z_{trans} and Z_{cis}. The overall stability of the different protonation states is controlled by the isomeric state of the chromophore.

Interestingly, other fluoroproteins contain a chromophoric moiety similar to asFP595, like, e.g., Dronpa [200], DsRed [256, 246], Kaede or KiKG [257, 258], eqFP611 [259, 260, 261, 262], Rtms5 [263], and HcRed [264]. In these structures *cis* or *trans* conformations of the chromophore have been observed. Our simulations on asFP595 suggest that chromophore photoisomerization could also be possible in these fluoroproteins. In particular, the similar structure of the reversibly photoswitchable protein Dronpa suggests that also the Dronpa chromophore can undergo *trans-cis* photoisomerization.

Chapter 6

Photostability of the Cytosine-Guanine DNA Base Pair

This project was done in close collaboration with Dr. Martial Boggio-Pasqua and Prof. Mike Robb from the Imperial College London. The static multiconfigurational calculations in the gas phase were carried out by Martial. Dr. Gerrit Groenhof carried out preliminary QM/MM excited state MD simulations to establish the validity of the applied methods. I carried out the excited state MD simulations and analyzed the data presented in Tables 6.2 and 6.5. The figures presented in this chapter were prepared together with Gerrit. The preparation of the DNA structure and the subsequent force field equilibration (prior to the QM/MM simulations) was done by Maik Götte.

6.1 Introduction

Deoxyribonucleic acid (DNA) carries the genetic information of all cellular forms of life. DNA usually forms a double helix, in which the nucleoside bases of the single strands are stacked upon each other, forming strong hydrogen bonds with the bases in the complementary strand (Watson-Crick configuration [81, 82]). Because the bases absorb light in the harmful ultra-violet (UV) region of the spectrum (wavelength ¡ 400 nm), DNA is vulnerable to photochemical damage. To protect the genetic information, highly elaborate mechanisms have evolved to repair damaged DNA [3, 4, 5]. More important, however, is the robustness of DNA with respect to UV damage. Indeed, when arranged in the Watson-Crick configuration, the isolated base pairs have an extremely short excited-state lifetime, suggesting a high photostability. Excited state decay measurements of bases and model base pairs suggest a subpicosecond repopulation of the ground state [49, 84]. Recent experiments by Abo-Riziq and co-workers indicate that the excited state of an isolated cytosine-guanine (C-G) base pair even has a lifetime in the order of a few tens of femtoseconds [83]. Recent static quantum chemistry calculations by Domcke and co-workers suggest that this ultra-fast deactivation may be triggered by a barrierless single proton transfer in the excited state [85, 86].

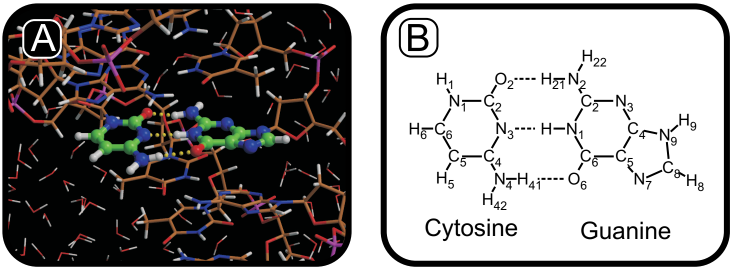


Figure 6.1: The cytosine-guanine base pair in the Watson-Crick configuration. **a)** Snapshot taken from one of the QM/MM molecular dynamics simulations of a 22 base pair DNA. The QM subsystem is shown in ball-and-stick representation; MM atoms are shown as sticks. Color code: carbon skeleton of the QM subsystem in green, MM carbon atoms in brown, nitrogen atoms in blue, phosphor atoms in magenta, oxygen atoms in red, and hydrogen atoms in white. **b)** Schematic representation of the base pair in vacuum, defining the atom names used in the text.

This chapter presents *ab initio* and QM/MM molecular dynamics simulations of a photoexcited cytosine-guanine (C-G) base pair in gas phase and embedded in DNA (Figure 6.1). The results of the simulations provide detailed structural and dynamical insights into the ultra-fast radiationless deactivation mechanism at a spatial and temporal resolution well beyond that of current experiments. According to the simulations, photon absorption to the charge transfer state (CT, Figure 6.3, a → d) promotes transfer of a proton from the guanine to the cytosine (Figure 6.2). Although in principle not actually essential for the decay, the proton transfer (PT) indirectly enhances ultra-fast radiationless deactivation of the excited state through an extended conical intersection (CI) seam. The conical intersection has an unusual topology, in that there is only *one* degeneracy-lifting coordinate. This topology is the central mechanistic feature for the decay in both vacuum and DNA. Radiationless decay occurs along an extended hyperline that lies nearly parallel to the proton transfer coordinate in the vicinity of the minimum of the charge transfer state (Figure 6.4), indicating that the proton transfer itself is not directly responsible for the deactivation. The seam is displaced from the minimum energy proton transfer path along a skeletal deformation of the bases. Decay can thus occur anywhere along the single proton transfer coordinate, accounting for the remarkably short lifetime of the excited base pair. In vacuum, decay occurs after a complete proton transfer, whereas in DNA, decay can also occur much earlier. The origin of this effect lies in the temporal electrostatic stabilization of the dipole in the charge transfer state in DNA.

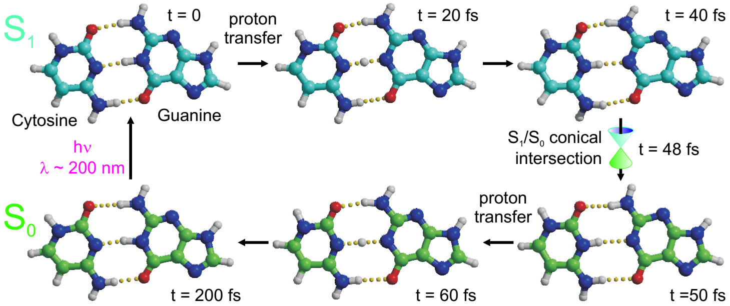


Figure 6.2: Photochemistry of the cytosine-guanine base pair in vacuum. Upon excitation to the CT state (S_1, cyan carbons) an electron transfer takes place from guanine to cytosine that is followed by a proton translocation (snapshot after 20 fs). The proton transfer leads to fast deactivation to S_0 (green carbons) through the conical intersection seam (48 fs). After returning to S_0, the proton is rapidly transferred back to the guanine base. The images are snapshots from one of our non-adiabatic molecular dynamics trajectories (run r, Table 6.2).

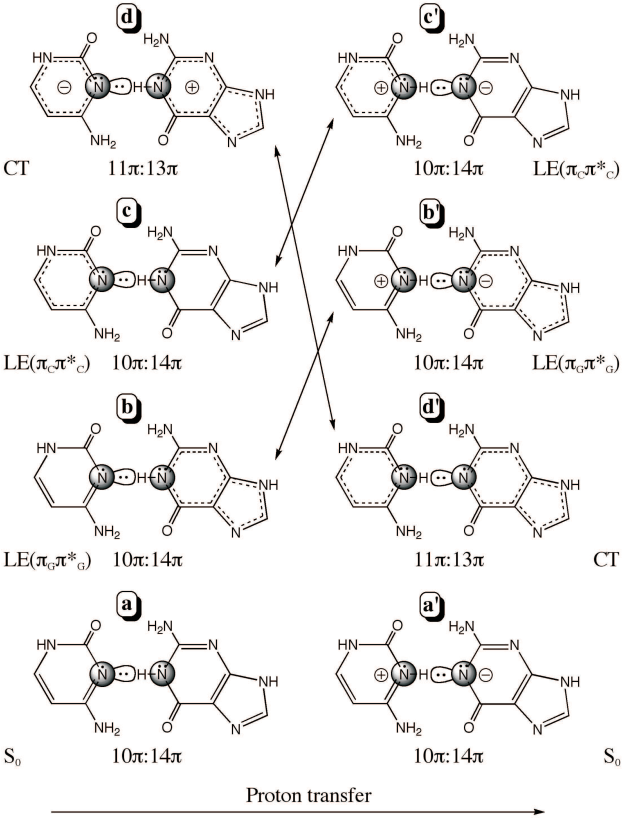


Figure 6.3: Valence bond representations of the relevant electronic states of the C-G base pair.

6.2 Methods

To model the dynamics of the photoactivated C-G base pair in our MD simulations, energies and gradients of both the excited CT (S_1) and ground states (S_0) were calculated on the fly at the CASSCF/3-21G level of theory [131, 132] with a reduced active space of 8 electrons in 8 orbitals. The reaction starts on the CT state but ends on S_0. The hops between the two electronic states were modeled by surface selection at the conical intersection (CI) seam. For all MD simulations Gromacs 3.3 [183] with an interface [63] to Gaussian03 [185] was used.

The rather small active space of 8 electrons in 8 orbitals was validated using geometry optimizations of relevant excited state minima and minimum energy crossing points for the isolated base pair. The full CASSCF active space for the C-G base pair would require 24 π electrons distributed in 19 π orbitals. Neglecting the lowest energy nodeless strongly occupied orbitals on each base and the corresponding virtual orbitals yields a manageable 20 electron, 15 orbital active space in which a single energy point calculation is feasible [265]. The final 8 electron, 8 orbital reduced active space was selected from a calculation using this larger active space to enable the simultaneous description of the electronic ground state and the lowest excited charge transfer state. The σ orbitals involved in the transfer of the central proton were not included within the active space because the proton transfer involves heterolytic bond cleavage.

At the Franck-Condon geometry, the CT state appears, in our calculations, as the lowest excited state (S_1) because of our particular selection of active space and orbitals, which biases this state relative to the locally excited (LE) states. Highly correlated CASPT2 computations by Domcke an coworkers [85, 86] show two LE states in between S_0 and the CT state (Figure 6.3, b and c). The relative energies of these states are within 40 kJ/mol of the CT state. These LE states would be initially populated upon excitation because of their much larger oscillator strengths. However, the LE states become higher in energy than the CT state upon after vibrational relaxation (see Figure 6 in reference [85, 86]). Thus, as we will discuss in the next section, the active space used in our MD simulations is a good representation of the CT state, and there is good agreement with the CASPT2 calculations by Domcke and co-workers (barrierless proton transfer pathway, CT minimum structure, and surface crossing with the ground state). Moreover, as Domcke and co-workers have also shown, there is virtually no barrier to reach a surface crossing, which justifies initiating our MD simulations directly in the CT state. Thus, we have not studied the vibrational relaxation taking place in one of the LE states, but rather we have focused on the single proton transfer channel only.

The starting coordinates for the QM/MM simulations of the fully hydrated 22 base pair B-DNA molecule were extracted from the crystal structure of the human DNA/topoisomerase I complex (entry 1A31 of the PDB [266]), in which the 5-iodo-2-deoxy-uracils were replaced by thymines and a break in the strand was fixed. As shown in Figure 6.1a, the C8-G37 base pair in the center of the molecule was described at the QM level, whereas the remainder of the system was modeled with the Amber99 force field [109]. The chemical bonds between the bases and the deoxyribose sugar rings connecting the QM and the MM subsystems were replaced by a constraint [226], and the QM part was capped with two hydrogen link atoms. The forces on the link atoms were distributed over the two atoms of the original bonds using the lever rule. The non-bonded interactions (Coulomb and Lennard-Jones) between the QM and the MM subsystems were calculated within a sphere of 1.6 nm. To avoid over-polarization of the QM subsystem, the charges on the MM atoms at the QM/MM boundary were set to zero, and the charges on the neighboring MM atoms were scaled to maintain an integer charge in the MM subsystem. In the QM/MM simulations, a time step of 0.5 fs was used, and no constraints were applied in the QM subsystem.

To equilibrate the DNA and the solvent prior to the QM/MM simulations, the system was simulated classically for 1000 ps. The simulation was performed in a rectangular $5.0 \times 4.3 \times 9.1 \text{nm}^3$ periodic box. After adding 6045 TIP4P water molecules [267], 42 sodium ions were inserted to compensate for the net negative charge of the DNA. The final system contained 25627 atoms. First, the system was energy-minimized (200 steps, steepest descent). Subsequently, a 250 ps MM simulation was performed with harmonic position restraints on all DNA heavy atoms (force constant $1000\,\text{kJ}\,\text{mol}^{-1}\text{nm}^{-2}$) for an initial equilibration of the water molecules and the ions. All simulations were run at constant temperature and pressure by coupling to an external bath [187] ($\tau_\text{T} = 0.1$ ps and $\tau_\text{p}= 1.0$ ps). The LINCS algorithm [226] was used to constrain bond lengths, allowing a time step of 2 fs in the classical simulations. SETTLE [188] was applied to constrain the internal degrees of freedom of the water molecules. A twin-range cut-off method was used for non-bonded interactions. Lennard-Jones interactions within 1.6 nm were calculated every time step, as were the Coulomb interactions within 1.0 nm. The particle mesh Ewald (PME) method [227] with a reciprocal grid spacing of 0.12 nm was used to calculate long range Coulomb interactions.

To generate the initial conditions for the excited state simulations, the systems (gas phase and condensed phase) were simulated for an additional 2 ps in the ground state at the CASSCF(8,8) level. From these ground state trajectories, frames at equal time intervals were used as starting structures for the excited state MD simulations.

6.3 Results and Discussion

Figure 6.4 shows a "cartoon" of the potential energy surfaces of the cytosine-guanine base pair for the ground (grey) and excited states (red) along with the path of a typical MD trajectory (dashed yellow/green line). The multi-dimensional surfaces are projected onto the proton transfer coordinate and the *single* degeneracy-lifting coordinate.

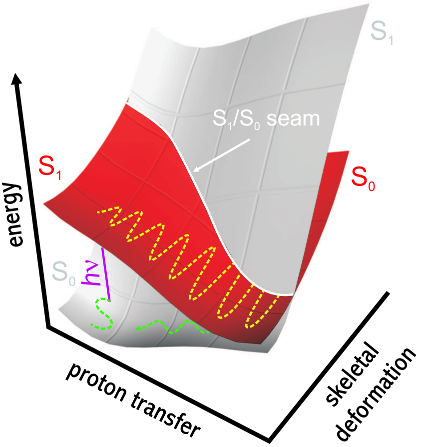


Figure 6.4: Schematic representation of the potential energy surfaces of the excited and ground states of the cytosine-guanine base pair versus proton transfer (N_1-H-N_3) and a skeletal deformation of the bonds (see Figure 6.7). The dashed yellow/green line represents the path sampled in a typical trajectory. Motion along the proton transfer coordinate connects the Franck-Condon geometry to the minimum of the CT state. Non-radiative decay occurs along the seam between the surfaces (conical intersection hyperline). The trajectory encounters the seam more than once and recrosses between CT and S_0.

The first coordinate involves transfer of the central proton ($N_1 - H \cdots N_3$, Figure 6.1b), whereas the latter describes a skeletal deformation of the base pair. The position of the S_1/S_0 intersection seam controls the passage of the trajectory from the excited state to the ground state.

6.3.1 Dynamics of the Isolated Base Pair

In the gas phase, 20 excited state simulations were initiated, each from a different frame of the 2 ps ground state trajectory. In these simulations, excited state lifetimes between 29 and 223 fs were observed (Table 6.2). Most trajectories showed several recrossings between the S_1 and S_0 surfaces (Table 6.2, 4th column). Such recrossing phenomenon has also been observed experimentally for the diatomic sodium iodide gas [268, 269]. On average three such recrossings were observed. Simple exponential fits yielded decay times of $\tau = 29 \pm 3$ fs and $\tau = 89 \pm 4$ fs until the first and final surface hop to S_0, respectively. These timescales favorably compare to the measurements of a fast decay time decay time of a few tens of femtoseconds [83]. The number of trajectories may seem small but nevertheless yields a consistent mechanistic picture.

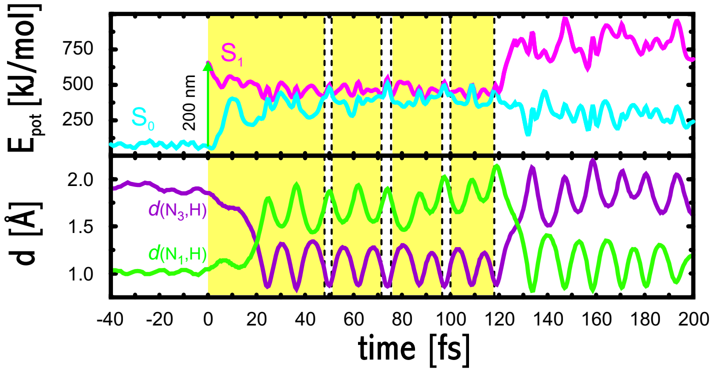


Figure 6.5: S_0 (blue) and S_1 (red) potential energies during a typical simulation (top panel, run h in Table 6.2); distances of the central proton to the donor (guanine N_1) and acceptor (cytosine N_3) (bottom). Surface hops are indicated by the dashed lines, time inervals on S_1 are colored yellow.

The excitation to the charge transfer state (CT), in which a single electron has been transfered from the guanine to the cytosine (a → d, Scheme 6.3), induces a spontaneous transfer of the central proton from guanine to cytosine within a few femtoseconds (Figure 6.2 and Scheme 6.3, d → d'). The bottom panel of Figure 6.5 shows the distance of this

run	$\tau_{(S_1)}$ 1st hop (fs)	$\tau_{(S_1)}$ last hop (fs)	rec.	$d_{\mathrm{N_3-H}}$ last hop (Å)
a	19	56	2	1.09
b	49	223	9	0.92
c	27	29	1	1.19
d	51	188	6	0.95
e	28	70	2	0.98
f	28	49	1	1.12
g	30	207	8	1.04
h	21	107	3	0.88
i	37	163	6	1.09
j	28	57	2	1.22
k	25	31	1	1.23
l	45	69	2	1.10
m	46	166	4	1.04
n	35	64	3	1.09
o	28	50	2	0.98
p	31	163	5	0.98
q	30	106	4	0.99
r	48	48	0	1.08
s	49	128	2	0.95
t	46	81	2	0.96

Table 6.2: Excited state lifetimes and number of surface hops in 20 simulations of the isolated cytosine-guanine base pair in vacuum. The two lifetimes in columns 2 and 3 refer to the time until the first and last surface hop to the ground state, respectively.

proton to its donor (guanine N_1) and acceptor (cytosine N_3) during a typical simulation. The last column of Table 6.2 lists the N_3-H distance at the time of the last hop, which ranges from 0.88–1.23 Å. In the CI minimum the N_3-H distance is 1.01 Å (Table 6.3), thus the hopping geometries are all close to that CI minimum. After the proton transfer, the system hits the conical intersection seam (Figure 6.2) and returns to the ground state, as indicated by the first dashed line. However, within a few femtoseconds, the seam is encountered again, and a second hop takes the system back to S_1, where it stays until another hop occurs. These recrossings are a manifestation of a specific CI surface topology that we will discuss below. Because of this topology, diabatic trapping is possible [270], and in principle the system could remain locked, switching back and forth between the S_1 and S_0 states and never reaching the S_0 minimum. However, this locking was not observed in any of our simulations, and after a finite number of such recrossings all trajectories eventually ended up on the ground state surface. On average the system spent 60 fs in this diabatic trap.

The hops to S_0 are accompanied by a shift of negative charge from the cytosine back to the guanine (d' → a', Scheme 6.3). This charge transfer provides the driving force for the central proton to return to the guanine after the final hop, which happens within a few femtoseconds (Figures 6.1 and 6.2, Scheme 6.3, a' → a).

In five out of twenty simulations the original Watson-Crick configuration was not restored, and the base pair eventually adopted a different tautomeric state, as shown in Figure 6.6.

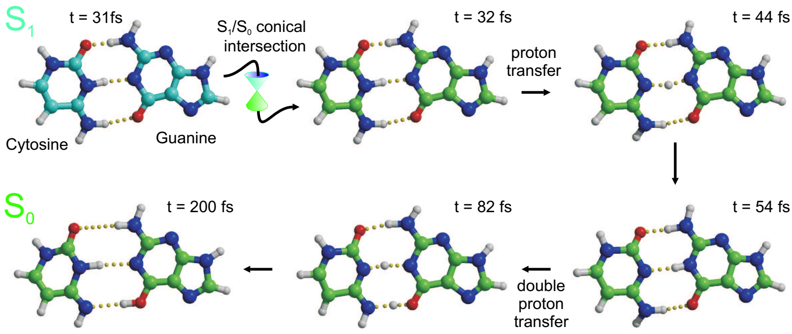


Figure 6.6: Formation of a different tautomer by a concerted double proton transfer. After deactivation of S_1 (cyan carbon atoms) through the conical intersection, the proton returns to the guanine. The excess thermal energy that is released upon returning to S_0 (green carbon atoms) induces a spontaneous double proton transfer (82 fs), which leads to the formation of a different tautomer (200 fs).

In three of these cases the central proton remained bound to the cytosine (N_3-H), and after the final surface hop to the ground state, a proton was transferred from the cytosine amino group (NH_2) to the guanine O_6 oxygen atom instead. In the two other cases the excess thermal energy released during the deactivation process induced a concerted double proton transfer in the ground state after the final radiationless transition, yielding the same tautomer. The system remained in this local minimum for the rest of the simulation, suggesting a barrier for re-tautomerization. This observation is in agreement with the experiments of Zewail and coworkers on model base pairs [271] and was also seen in recent CPMD simulations [272]. In solvated double stranded DNA, where the bases are held tightly together by the phosphate backbone, the excess thermal energy can be channeled away more efficiently, and such potentially mutagenic double proton transfer events were not observed (see below).

6.3.2 S_1/S_0 Intersection Topology

This subsection discusses the conical intersection that occurs between a' and d' (Figure 6.3) in the region of the CT minimum. Photon absorption triggers a transition from the covalent ground state (Scheme 6.3, a) with 10 and 14 electrons in the π systems of the cytosine and guanine, respectively, to the zwitterionic CT state (Scheme 6.3, d) with 11 and 13 π electrons, respectively. This CT state is diabatically correlated to the non-zwitterionic CT minimum with the central proton transferred to the cytosine (Scheme 6.3, d').

The excited state CT minimum and an adjacent minimum energy conical intersection were optimized. Table 6.3 lists the bond lengths at these critical points.

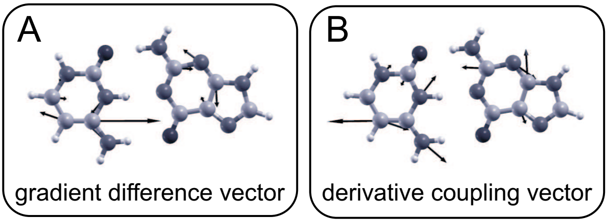


Figure 6.7: The computed gradient difference (a) and derivative coupling vectors (b) at the optimized conical intersection geometry, which is almost identical to the S_1 minimum (Table 6.3). The plotted amplitudes are arbitrary. The vectors in (a) and (b) are obviously co-linear and of opposite phase. Thus they are linearly dependent and span a one-dimensional branching space.

bond	CT min	CI min	bond	CT min	CI min
N1-C2	1.358	1.356	N1-C2	1.307	1.307
C2-O2	1.238	1.238	C2-N2	1.317	1.317
C2-N3	1.353	1.354	C2-N3	1.414	1.416
N3-H	1.011	1.010	N3-C4	1.274	1.273
N3-C4	1.412	1.415	C4-C5	1.442	1.441
C4-N4	1.373	1.377	C5-C6	1.439	1.441
C4-N5	1.368	1.365	C6-O6	1.247	1.246
C5-C6	1.402	1.405	C6-N1	1.381	1.383
C6-N1	1.419	1.419	C5-N7	1.368	1.366
			N7-C8	1.325	1.326
			C8-N9	1.398	1.397
			C4-N9	1.367	1.368

Table 6.3: Bond lengths (in Å) of the cytosine (left) and the guanine (right) bases at the optimized minima of the charge transfer state and the conical intersection seam.

The central proton is completely transferred to the cytosine in these configurations, with the same $N_3 - H$ bond distance of 1.01 Å in the CT minimum and at the conical intersection (Table 6.3). This is in good agreement with Domcke's highly correlated calculations [85, 86], which locate the CT minimum at an $N_1 - H$ bond distance of 2.0 Å, compared to 1.98 Å in our calculations. In the isolated base pair we found the CI minimum to lie only 1 kJ/mol above the CT minimum. Thus, within the accuracy of the applied methods, the CT minimum and the CI minimum have about the same energy. The same energies indicate that the seam is encountered after adiabatic reaction on the CT state surface. Other minima or other topological features between the Franck-Condon region and the CT minimum were not found.

At the conical intersection, the gradients of the ground and CT state are parallel, indicating that the CI is sloped. The gradient difference vector and the interstate coupling vector that span the branching space of the conical intersection are co-linear and correspond to a skeletal deformation of the two bases (Figure 6.7). These vectors do not contain a first order component of the proton transfer. The conical intersection therefore lies on a hyperline that is parallel to the proton tranfer coordinate (Figure 6.4). A quadratic analysis of the intersection seam revealed that the coupling to proton transfer is of second order and that the seam curves along this coordinate.

Because the derivative coupling and gradient difference vectors that lift the S_1/S_0 degeneracy of the seam are co-linear, the branching space is one-dimensional. The conical intersection hyperline thus has a dimension of (N^{int}-1) instead of (N^{int}-2), where N^{int} is the number of internal degrees of freedom of the system (see Section 2.3.3). The unusual character of this (N^{int}-1) dimensional sloped intersection seam means that it cannot be avoided, creating a diabatic trap for the CT state.

In summary, the S_1/S_0 intersection seam in the base pair is sloped and there is only *one* coordinate that lifts the degeneracy. This topology accounts for the recrossings (partial diabatic trapping) that were observed in the dynamic simulations. As schematically shown in Figure 6.4, vibrational motions along this coordinate repeatedly drive the system in-and-out of the CI, until the base pair eventually returns to its stable ground state minimum. The proton transfer coordinate is perpendicular to the one-dimensional branching space and therefore does not lift the degeneracy. The seam is accessible anywhere along the proton transfer coordinate, and in principle proton transfer is not essential for the S_1 decay. Rather, since in the excited CT state proton transfer is a barrierless downhill process, part of the excitation energy is rapidly channeled into excess vibrational energy of the other degrees of freedom (through intra-molecular vibrational relaxation), including the skeletal deformation that leads to the intersection seam. Therefore, the proton transfer indirectly facilitates the ultra-fast radiationless decay.

6.3.3 Dynamics of the C-G Base Pair in DNA

Thus far, this chapter described the mechanism that underlies the ultra-fast excited state decay process of an isolated base pair. To study how the DNA surrounding influences this process, we have performed 20 excited state QM/MM MD simulations of a cytosine-guanine base pair embedded in a fully hydrated double stranded B-DNA molecule (Figure 6.1a). Also in DNA, the excited state rapidly decays through the conical intersection seam. The average excited state lifetimes of $\tau = 48 \pm 4$ fs and $\tau = 137 \pm 6$ fs until the first and last surface hop, respectively (Table 6.5) are somewhat longer than in vacuum. However, there is no fundamental difference between the processes in vacuo and in condensed phase, which can both be considered ultra-fast excited state decay processes. An average of four S_1/S_0 recrossings were observed in the DNA, one more than

run	$\tau_{(S_1)}$ 1st hop (fs)	$\tau_{(S_1)}$ last hop (fs)	rec.	$d_{\mathrm{N_3-H}}$ last hop (Å)
a	26	352	12	0.99
b	27	87	3	0.94
c	28	102	5	1.35
d	26	26	0	1.15
e	30	98	3	1.04
f	28	138	1	1.05
g	58	172	10	1.09
h	174	176	1	0.94
i	181	205	1	0.88
j	30	211	4	1.15
k	12	101	2	1.00
l	11	11	0	1.89
m	23	197	7	1.04
n	27	128	6	0.90
o	25	68	1	1.07
p	67	498	13	1.08
q	122	259	6	1.19
r	30	84	2	1.00
s	11	87	2	1.14
t	145	145	0	1.83

Table 6.5: Excited state lifetimes and number of surface hops in 20 simulations of a cytosine-guanine base pair in hydrated DNA. The two lifetimes in columns 2 and 3 refer to the time until the first and last surface hop to the ground state, respectively.

in vacuum. The $N_3 - H$ distances at the final crossings (last column) ranged between 0.88 Å and 1.89 Å, spanning the full extent of the CI seam.

The N_3-H distances at the last hop can be divided in two sub populations. Most of the hops occur near the CT minimum, but in two out of the twenty DNA trajectories the S_1/S_0 seam was hit before proton transfer could occur (runs l and t, Table 6.5). In these cases no recrossings were observed, which is a natural consequence of the extended seam (Figure 6.4). However, this part of the seam is not accessible in the gas phase. These observations suggest that the DNA environment can temporarily lower the energy of the charge transfer state and thus bring the crossing seam closer to the Franck-Condon region than in vacuum. This conjecture was tested by computing the overall electrostatic stabilization of a CT dipole model that was created by placing a point charge in the center of each six-membered ring of the base pair. The total electrostatic potential was then evaluated at every frame of the trajectories. In runs l and t, the CT dipole is stabilized by the DNA environment by around 100 kJ/mol.

In DNA, the base pair always returned to the initial Watson-Crick configuration after deactivation, and no double proton transfer events were observed. The interactions with the rest of the DNA kept the base pairs tightly together. In this respect the base pair is more photostable in DNA than in vacuum, where the recovery of the original ground state Watson-Crick configuration was only observed in 75% of the cases. In the other 25% a wrong tautomer was formed.

By including only a single base pair in the QM region, the possibility of observing possible intra-strand excimer formation was excluded. Recent experiments by Kohler and co-workers on an $(AT)_{18}$ B-DNA molecule have revealed that excited state relaxation mainly involves such intra-strand excimer formation [273]. The measured excited state lifetime was about 50-150 ps, indicating that the ultra-fast proton transfer is not the predominant decay channel in $(AT)_{18}$. This observation suggests that also for DNA containing C-G base pairs, such excimer formation could play an important role in the decay process. We must assume that there could be a competition between entering the charge transfer state of the base pair and the excimer formation. The efficiency of the first process is strongly dependent on the hydrogen bonding configuration, as Domcke and co-workers have shown [85, 86]. All our QM/MM simulations were initiated from ideal Watson-Crick configurations, with three hydrogen bonds between the bases. These configurations were by far more predominant during a 50 ns classical ground state DNA simulation than partially unpaired base pairs. Entering the CT state from such partially unpaired configurations could be more difficult, leading to a different decay process. Studying alternative pathways such as excimer formation is a formidable task, and is beyond the scope of this work.

6.4 Conclusions

Understanding the dynamics of the ultra-fast deactivation processes in the cytosine-guanine base pair represents a major challenge. This chapter has shown that by using an *ab initio* MD strategy with explicit surface hopping, it is not only possible to reproduce the experimental excited state lifetime of the isolated cytosine-guanine base pair, but also to provide detailed mechanistic insights into the deactivation process. The important mechanistic aspect of the CI surface topology is the presence of an extended seam parallel to the reaction path, so that decay can occur in principle anywhere along the proton transfer coordinate *via* a skeletal deformation. In the gas phase the reaction path is sufficiently steep along the proton transfer coordinate for the system to become partially equilibrated in the charge transfer state minimum before radiationless decay. In the DNA, this process appears to be prevalent as well. However, decay was also observed occasionally near the Franck-Condon region due to electrostatic stabilization of the CT resonance structure by the DNA surroundings. We expect that this work will stimulate further experimental and theoretical studies to validate our conclusions, in particular, the ultra-fast decay time and the predicted oscillatory repopulations of the excited state.

Chapter 7

Elastic Properties of Azobenzene Polymers

7.1 Introduction

Nanomechanical devices or molecular machines will, for a broad range of applications, most likely be powered by light or other kinds of electromagnetic radiation [10, 11, 12, 13]. The major reasons are ease of addressability, picosecond reaction times to external stimuli, and compatibility with a broad range of ambient substances, such as solvents, electrolytes, or gases. Azobenzene is a well-studied photoactive system, which can be photoswitched selectively from an extended *trans* and a more compact *cis* conformation using light of wavelength 365 nm, whereas the reverse *cis*-to-*trans* isomerization is induced by light of wavelength 420 nm (Figure 7.1a) [87, 88]. Many processes, such as light-driven ion transport through biological membranes [274, 275, 276], can be steered by conformational switching of azobenzene chromophores. Azobenzene has been used frequently in synthetic photoresponsive systems for the regulation of the geometry and function of biomolecules [277, 278, 279, 280, 281, 282, 283, 35, 284, 285].

The isomerization of individual azobenzenes has also been studied by scanning tunneling microscopy [286, 287]. Recently, a bistable polyazobenzene peptide was synthesized as a model system for a light-powered molecular machine, and its mechanical properties were characterized by means of single-molecule atomic force microscope (AFM) experiments

(Figure 7.1b, c) [89, 90]. Optical switching of the azobenzene polymers between their extended *trans* and compact *cis* conformations was demonstrated, and the corresponding change in the contour length of the polymer was detected. Thereby, in analogy to an Otto cycle, Gaub and co-workers established an optomechanical operating cycle, in which optical contraction against an external force delivered net mechanical work [89, 90]. Thus, they demonstrated that azobenzene polymers indeed hold great promise for future applications in nanotechnology, for example, as light-triggered molecular switches or cargo lifters. In a related experiment, Vancso and co-workers characterized a redox-mechanical cycle by using electrochemical AFM-based single-molecule force spectroscopy [288].

However, the measured overall length change upon switching between the *cis* and *trans* polymers — and thus the work output — was considerably smaller than could be expected on the basis of the length change of a single azobenzene monomer and the number of individual monomers. To resolve this discrepancy, and, more generally, to obtain a detailed microscopic understanding of the underlying polymer mechanics at the atomistic scale, we carried out explicit solvent force-probe molecular dynamics (FPMD) simulations [165, 166] of polyazobenzene model peptides under mechanical stress, the results of which resemble those of the AFM experiments described in Figure 7.1c. The simulations provided detailed mechanistic insight that is prerequisite for the efficient and targeted optimization of photoswitchable polymers for future applications in nanotechnology. Because we were only interested in the elastic properties of these model polymers, and not in the photoisomerization kinetics, we kept the azobenzenes fixed in the *cis* or *trans* conformation during the simulations. We did not include the photoisomerization itself into our simulations on the assumption that this process occurs on much faster time scales than the subsequent structural rearrangement of the polymer. Indeed, sub-picosecond kinetics have been observed for azobenzene photoisomerization [289, 290] which, is a complex process involving several electronically excited states [291, 292, 289, 293, 294, 295, 290].

By comparing our simulations to the AFM experiments, we gleaned insight into the overall elastic properties of azobenzene polymers and could relate our results directly to the measured elasticity characteristics. We were able to elucidate the crucial role of the peptides that interlink the azobenzene units and to predict the elastic behavior beyond the experimentally accessible force regime. Finally, we used the detailed mechanistic understanding thus obtained to design a photoswitchable polyazobenzene peptide with improved optomechanical properties.

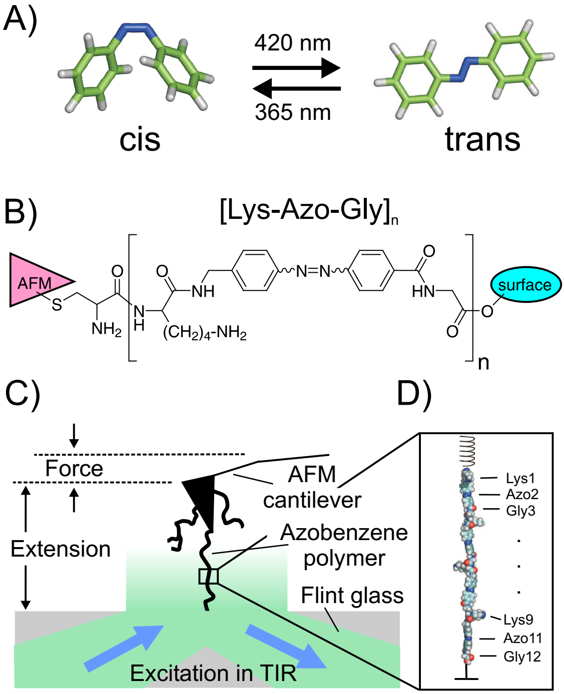


Figure 7.1: **a)** Photochromic azobenzene can be switched between a compact *cis* and an extended *trans* conformation. **b)** A polyazobenzene peptide composed of Lys-Azo-Gly azotripeptide units. **c)** Sketch of the AFM experiment which motivated the study described herein [89, 90]. The azobenzene polymer was attached to the AFM tip through a cysteine linker (top), and mounted to an amino-functionalized surface (bottom); adopted with permission of the authors of ref. [89]. **d)** Space-filling representation of the all-*trans* conformation of the model polymer used for the force-probe MD simulations. To mimic the AFM experiments, a harmonic spring was attached to the N-terminus (top) and pulled upwards with constant velocity, while the C-terminus was kept fixed (bottom).

7.2 Simulation Details

7.2.1 Setup

The model azobenzene polymers used in our simulations were constructed from four azotripeptide units ($n = 4$ in Figure 7.1b). These (Lys-Azo-Gly)$_4$ dodecamers have the same stoichiometry as the polymers used in the AFM studies. As in in the experiments, DMSO (dimethylsulfoxide) was used as a solvent. To study the effect of the azobenzene conformation (*cis* or *trans*) on the elastic properties, we considered three stereochemically different polymers, an all-*cis*, an all-*trans*, and a mixed *cis-trans-cis-trans* system. The mixed polymer was considered because complete photoisomerization upon optical pumping was not achieved in the experiments as a result of the overlapping absorption bands of the *cis* and *trans* states [89, 90]. The study of this polymer would also reveal possible cooperative effects between adjacent monomers.

All simulations were carried out with the Gromacs simulation suite [183], using the OPLS all-atom force field [108] and periodic boundary conditions. NpT ensembles were simulated with the polymer and solvent coupled separately to a 300 K heat bath ($\tau_\mathrm{T} = 0.1\,\mathrm{ps}$) [187]. The systems were isotropically coupled to a pressure bath at 1 bar ($\tau_\mathrm{p} = 1.0\,\mathrm{ps}$) [187]. Application of the Lincs [226] algorithm allowed for an integration time step of 2 fs. Short-range electrostatic and Lennard-Jones interactions were calculated within a cutoff of 1.0 nm and 1.4 nm, respectively, and the neighbor list was updated every 10 steps. The particle mesh Ewald (PME) method was used for the long-range electrostatic interactions [227]. To compensate for the net positive charge of the protonated polymers, four chloride ions were added.

The OPLS force constants of the angles involving the central azo-moiety are crucial for the elastic properties and were determined using quantum chemical calculations. Relaxed potential energy surface scans along a stretching coordinate defined by the distance between the azobenzene *p*-carbon atoms were performed at the B3LYP/6-311+G* density functional level [221, 296] using Gaussian03 [185]. As shown in Figure 7.2, the OPLS parameters were chosen such as to accurately describe the potential energy along the stretching coordinate as well as the corresponding structures. For the rotation around the central N=N bond, a dihedral potential barrier of 100 kJ/mol was chosen, in accordance with the measured barrier of 108 KJ/mol [297]. Atomic partial charges of azobenzene were obtained from B3LYP/6-31G* calculations according to the CHELPG scheme [228]. The employed force field parameters are summarized in the Tables in the Appendix.

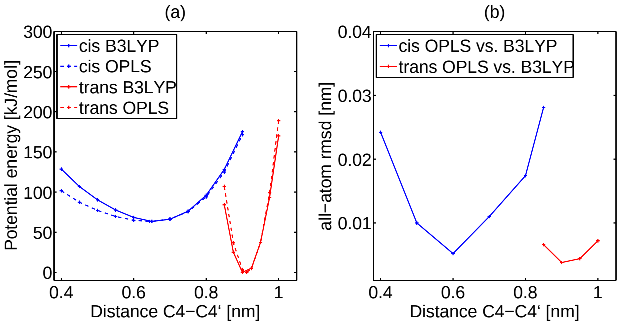


Figure 7.2: Parametrization of the OPLS force field for azobenzene against B3LYP/6-311+G* density functional calculations. **a)** Potential energy along the C4-C4' stretching coordinate of azobenzene in vacuo. The curves for *cis* and *trans* azobenzene are shown in blue and red, respectively. The OPLS force field was parametrized such that the force field potential energy (dashed lines) closely matches the B3LYP/6-311+G* energy (solid lines). The energy offset between the *cis* and *trans* conformers was taken from the DFT calculation. Note that *cis* azobenzene is significantly softer as compared to *trans*. **b)** All-atom rmsd between the force field and DFT structures along the stretching coordinate.

The starting configurations for the FPMD simulations were obtained as follows. First, the azobenzene polymers were constructed using pymol [298]. In these structures, the lysines and glycines were modeled in their extended anti conformations, with ψ and φ backbone dihedral angles of 135° and $-140°$, respectively. Second, the polymers were solvated in a cubic box of DMSO and energy minimized (1000 steps steepest descent), followed by 500 ps of MD with positional restraints on all polymer heavy atoms to equilibrate the solvent. Then, the whole system was equilibrated for 1 ns, after which the solvent was removed. The polymer was aligned along the z-axis, and the simulation box was extended in the z-direction to enable the accommodation of the fully stretched conformation. Finally, DMSO was added to this elongated box and equilibrated with positional restraints on all polymer heavy atoms for another 500 ps to yield the starting configurations for the FPMD simulations. Depending on the capping of the lysine side chains and the azobenzene conformations, total system sizes were between 35,000 and 57,000 atoms.

7.2.2 Force Probe MD Simulations

During the force-probe simulations, the N-terminus N_α was subjected to a moving harmonic spring potential (eq 2.51 in Section 2.5) with a spring constant of $k_0 = 500\,\mathrm{kJ\ mol^{-1}nm^{-2}}$. The C-terminus C_ω was fixed by a stationary harmonic potential ($k_{fix} = 1000\,\mathrm{kJ\ mol^{-1}nm^{-2}}$). Mechanical stress to probe the elastic behavior of the model polymers was applied by moving the spring with constant velocity v in positive z-direction, $z_{spring}(t) = z(0) + vt$, with $z(0) := z_{N_\alpha}(0)$. During the FPMD simulations, z_{spring}, z_{N_α}, and the position of the C-terminus were recorded every 0.1 ps, whereas the positions of all polymer atoms were recorded every 0.5 ps. For further analysis, the force at the spring potential (calculated using eq 2.52) was averaged using a 50 ps time window. The data for the wormlike-chain (WLC) fits were truncated at 500 pN, which was the maximal force in the AFM experiments. The FPMD simulations were carried out with pulling velocities of $v = 0.1\,\mathrm{nm/ns}$ and $v = 0.5\,\mathrm{nm/ns}$ to yield simulation times of 50–70 and 10–14 ns, respectively, per trajectory. The total simulation time was about 500 ns.

7.3 Results and Discussion

7.3.1 Elastic Properties

The force-versus-extension traces obtained from the AFM experiments reflect the elastic properties expected for a polymer subjected to a mechanical load. Figure 7.3a shows the overall force-extension traces of the all-*cis*, all-*trans*, and mixed azobenzene model polymers obtained from our force-probe MD simulations (over 50–70 ns each). These traces closely resemble the curves from the AFM experiments (Figure 7.3a, inset) [89, 90]. At higher forces above 600 pN, deviations from the WLC behavior are seen. These deviations are discussed below. As expected, the force trace of the mixed *cis-trans-cis-trans* polymer lies midway between those of the all-*cis* and the all-*trans* polymers. This result already suggests that cooperative effects between monomers are unlikely. Therefore, such effects cannot explain the unexpectedly small length changes observed in the AFM experiments.

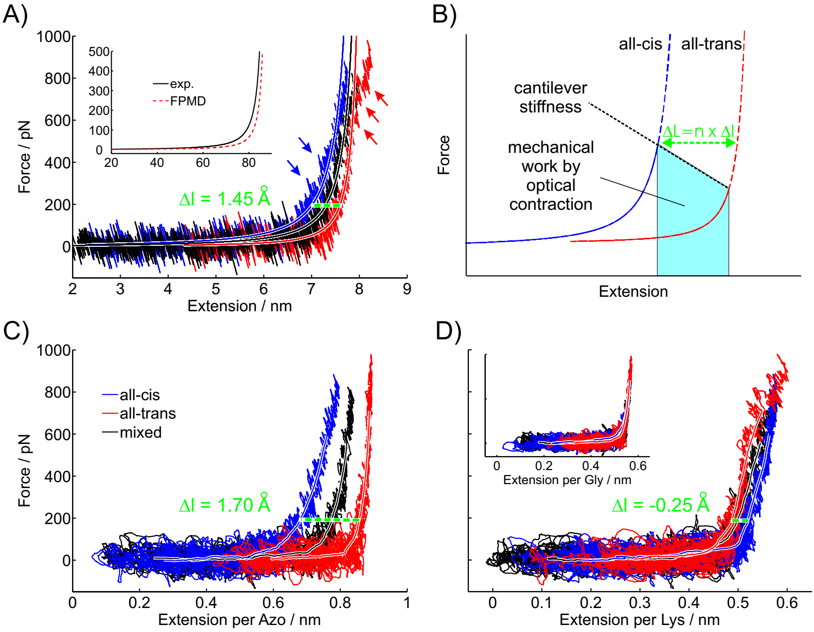


Figure 7.3: Force-probe MD force-extension curves for (Lys-Azo-Gly)$_4$, obtained at a pulling velocity of 0.1 nm/ns. **a)** Curves of the all-*cis* (blue), all-*trans* (red), and mixed *cis-trans-cis-trans* (black) polymers. The overall extension is defined as the distance between the C-terminus and the N-terminus along the pulling direction. The extension difference obtained from a WLC fit (smooth lines) of 1.45 Å per unit between the *cis* and *trans* polymers (at 200 pN) is shown as a green dashed line. Arrows indicate events discussed in the text. Inset: WLC curves obtained from experiment [90] (black line) and simulation (red dashed line) for the all-*trans* polymer, for the contour length of $L_c = 89.1\,\mathrm{nm}$ taken from ref. [90]. **b)** The output of mechanical work (blue trapezoid) due to optical contraction against an external load increases with the extension difference, ΔL, which is the extension difference per monomer, Δl, multiplied by the number of monomers, n . The cantilever stiffness is reflected in the slope of the dashed black line. **c,d)** Force-extension curves of the individual polymer building blocks azobenzene (c), lysine (d), and glycine (d, inset). The smooth lines represent 100,000-point averages.

The force field and simulation protocol used were validated by comparison with the AFM results. To this end, the force-extension data obtained from our simulations were fitted to a WLC model [299],

$$F(r) = \frac{k_{\mathrm{B}}T}{L_{\mathrm{p}}}\left[\frac{1}{4}\left(1-\frac{r}{L_{\mathrm{c}}}\right)^{-2}-\frac{1}{4}+\frac{r}{L_{\mathrm{c}}}\right], \qquad (7.1)$$

in which F is the force, $T = 300\,\mathrm{K}$ the temperature, r the polymer extension, L_{c} the contour length, and L_{p} the persistence length. Table 7.3.4 lists the WLC parameters obtained as well as the extension difference per switched azobenzene monomer, Δl, for all simulated systems, together with the values obtained from experiment. Because experimental WLC data are only accessible for the all-*trans* polymer (Figure 4 in ref. [90]), we compare in Table 7.3.4 only the parameters of the all-*trans* polymers. The contour length L_{c} per azotripeptide unit given in ref. [89] was not extracted from the measured force traces, but determined by molecular modeling. Therefore, we only compare the persistence lengths L_{p}. These lengths obtained from the simulations range from 0.40 nm to 0.97 nm and agree with the experimental value of 0.5 nm (Table 7.3.4) within our statistical accuracy. We found L_{p} to be quite sensitive to the details of the fit. The quality of the fit did not drop significantly when the experimental value of $L_{\mathrm{p}} = 0.5\,\mathrm{nm}$ was used (R^2_{fix} in Table 7.3.4). This good agreement between theory and experiment validates the force field applied and renders finite size effects of the dodecamers used as model systems for the more extended azobenzene polymers used in the AFM experiments unlikely. Further tests to confirm the validity of our approach are presented in Section 7.3.4.

Because the chemistry of the lysine side chains of the polymers used in the AFM experiments could not be unambiguously established, we repeated all simulations with (i) all lysine side chains charged, and (ii) all lysine side chains capped with an adamanyl-oxycarbonyl (Adoc) protection group. Hence, taken together with the uncharged model polymer described above, the lysine side chains were modeled in all conceivable ways. The WLC parameters (Table 7.3.4) show that the elastic properties of the charged and Adoc-capped model systems do not differ significantly from the uncharged model polymer. As we will detail below, this is because the elastic behavior is dominated by the polymer backbone, and not by the side chains.

The WLC fits were also used to extract the extension differences Δl between the *trans* and *cis* polymers from the simulations. As illustrated

system	Δl (Å)	L_c (nm)	L_p (nm)	R^2_free	R^2_fix
exp. (ref. [89])	0.60	1.90	0.50		
uncharged (sim 1)	1.45	2.05	0.97	0.88	0.85
uncharged (sim 2)	1.48	2.07	0.68	0.92	0.91
uncharged (sim 3)	1.49	2.12	0.40	0.89	0.88
charged	1.34	2.11	0.41	0.92	0.92
Adoc-capped	1.38	2.10	0.45	0.87	0.87

Table 7.1: WLC parameters were obtained by modeling the force-extension data with the extended WLC model (Eq. 7.1). The extension difference Δl per switched azobenzene monomer is given at a force of 200 pN to enable direct comparison with the experimental data. Two different values are given for R^2; these values reflect the quality of the fits. R^2_free was obtained from a WLC fit with both L_c and L_p as adjustable parameters, whereas R^2_fix was obtained from a fit with $L_\mathrm{p} = 0.5\,\mathrm{nm}$, which is the value obtained from the experimental data. For the uncharged model polymer, Force-Probe MD simulations were carried out at pulling velocities of 0.1 nm/ns (sim 1) and at 0.5 nm/ns (sim 2 and sim 3).

in Figure 7.3b, Δl determines the work output upon contraction against an external load. Therefore, an increase in Δl improves the efficiency of the optomechanical energy conversion caused by the polymers.

The extension difference between the *cis* and *trans* states of an isolated azobenzene molecule was estimated from quantum chemical calculations to be 2.5 Å [291]. However, the polymer typically does not undergo the corresponding length change, because at low forces ($F < 200\,\mathrm{pN}$) the azobenzene monomers are not aligned perfectly, and thus only part of the *cis-trans* length change of the monomer is available in the pulling direction. In contrast, at higher forces the polymers are aligned to a larger extent, and thus could undergo larger length changes. However, *trans* azobenzene is considerably stiffer than *cis* azobenzene (Figure 7.2a), which again reduces the achieved length change. From our simulations, in which these effects were taken fully into account, we predict an extension difference of about 1.3 Å to 1.5 Å per switched azobenzene. As the measured extension difference of about 0.6 Å is still smaller by more than a factor of two (Table 7.3.4), the work output might obviously be improved. Closer inspection shows that only part of this reduction in polymer length change Δl can be explained by the above two effects.

7.3.2 Contributions of Individual Residues

To identify the origin of the as yet unexplained reduction in Δl, we decomposed the overall force-extension curves into the contributions of the individual polymer building blocks. Indeed, the extension difference solely due to the azobenzenes is 1.70 Å per monomer (dashed green line Figure 7.3c), which is the value expected from polymer tilt and differential stiffness. This contribution to Δl is larger than the total extension difference of 1.45 Å for the polymer (Figure 7.3a). Figure 7.3d shows that the difference of about 0.25 Å can be attributed to the lysine residues, which therefore partly compensate the optical lengthening and contraction of the azo groups. The force-extension curves of the glycine residues are rather independent of the azobenzene conformation (Figure 7.3d, inset).

As schematically shown at the right-hand side of Figure 7.4d, the compensation by the lysines is due to the isomerization of the lysine backbone between an extended *anti* and a more compact *syn* conformation around the backbone dihedral angle ψ. These two conformers are populated to different extents, depending on whether the adjacent azobenzene moiety is *cis* or *trans* (Figure 7.4). Figure 7.4b,c shows exemplarily the backbone dihedral angles ψ of Lys7 in the all-*cis* and in the all-*trans* polymers, respectively, as a function of the applied force (black), together with the average of the ψ angle and the statistical error (blue and red for *cis* and *trans*, respectively). In Figure 7.4d, the average of the backbone dihedral angle ψ of all lysine residues is shown as a function of the applied force. For the all-*cis* polymer, the lysines already adopted the extended *anti* conformation at very low forces between 0 and 300 pN (blue lines). These transitions reveal themselves also as steps in the force-extension trace of the all-*cis* model polymer (blue arrows in Figure 7.3).

In the mixed polymer, Lys7, which is directly bonded to a *cis* azobenzene (Figure 7.4a) also adopted the *anti* conformation very early (dash-dotted black line in Figure 7.4d), in a similar manner to Lys4 in the all-*cis* polymer. In contrast, Lys4 and Lys10 with adjacent *trans* azobenzene units underwent *syn*-to-*anti* transitions only at rather large forces between 600 and 800 pN (solid and dashed black lines, respectively). The same behavior was observed for all lysine amino acids in the all-*trans* polymer (red lines).

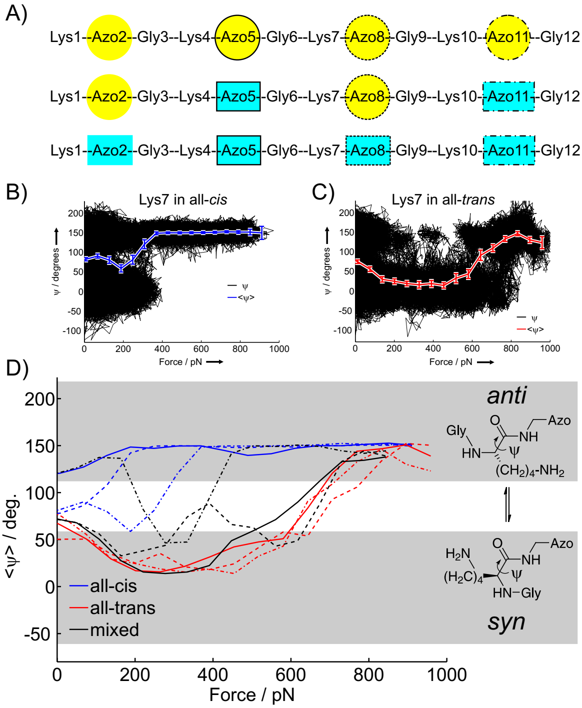


Figure 7.4: *Syn*-*anti* conformational switching of lysine amino acids. **a)** Sketch of the all-*cis* (upper), mixed (middle), and all-*trans* (lower) polymers. Azobenzenes in *cis* conformation are shown as yellow circles, those in *trans* conformation as cyan rectangles. **b,c)** Exemplarily, the backbone dihedral angles ψ of Lys7 in the all-*cis* and all-*trans* polymers are plotted as a function of the applied force. Averages of ψ were calculated using a 50 pN window. Error bars represent the errors of the mean and are typical also for the data shown in panel d. (d, right) Extended anti and compact syn lysine conformations interconvert via rotation around the backbone dihedral angle ψ. **d)** Averaged ψ angles of all individual lysine monomers are plotted as a function of the applied force for the all-*cis* (blue), all-*trans* (red), and mixed (black) polymers for Lys4 (solid), Lys7 (dash-dotted), and Lys10 (dashed line). All forces were obtained at a pulling velocity of 0.1 nm/ns. The gray-shaded areas indicate the extended anti and the more compact syn conformations. Lys1 is directly attached to the harmonic pulling potential (Figure 7.1d) and was therefore excluded from the analysis.

In summary, our simulations showed that lysine residues bonded to *trans* azobenzene moieties tend to adopt the compact *syn* conformation (up to 600–800 pN), whereas lysines adjacent to *cis* azo groups undergo *syn*-to-*anti* conformational transitions already at low to moderate forces below 300 pN. This preference of the extended *anti* conformation of lysines adjacent to *cis* azobenzenes explains the observed compensation effect. The difference in the length of the polymer chain due to *cis-trans* isomerization of the azobenzene units is reduced from 1.7 Å to 1.3–1.5 Å. However, in the AFM experiments, a value of only 0.6 Å was measured. We conclude that in the AFM experiments only about 40 % of the azobenzenes were effectively photoisomerized upon optical pumping. The major reason for the failure to attain the maximum value, which is about 70–80 % as a result of the overlapping absorption bands of the cis and *trans* isomers [300], in the AFM experiments is probably that only about half of the polymer was excited by the evanescent light field, which agrees with the estimate of the authors [89, 90, 301].

7.3.3 Predictions

The simulations also allow the investigation of the polymer mechanics at high forces and thus enable prediction of the elastic behavior even well beyond the force regime accessible to AFM. Only at forces larger than 1000 pN were mechanically induced *cis*-to-*trans* transitions of the azobenzene moieties observed in our simulations (data not shown). These isomerizations occurred by rotation around the central N=N bond, in agreement with recent *ab initio* calculations [301]. The high forces observed are consistent with the large thermal barrier to isomerization of 108 kJ/mol [297], which, as a result of the orthogonality of the rotation coordinate with respect to the pulling coordinate, is largely unaffected by the applied force [301]. Also in the AFM experiments no mechanical transitions were observed up to 500 pN [89, 90]. At large forces of about 600–800 pN, the simulations also predict *syn*-to-*anti* transitions of the lysine residues in the all-*trans* polymer. For the short polymer used in this study, these transitions are visible as steps in the force-extension curve (red arrows in Figure 7.3a). For the much larger polymers used in the AFM experiment, these individual steps would not be be resolved, but a tendency towards larger extensions would be observed in the force-extension traces of the *trans* polymers at forces larger than about 600 pN.

7.3.4 Controls

To test the influence of the pulling velocity on the elastic properties, that is, to check whether the system was sufficiently close to equilibrium during the simulations, we carried out two additional simulations of the uncharged (Lys-Azo-Gly)$_4$ dodecamer at a pulling velocity of 0.5 nm/ns. The overall and individual force-extension curves as well as the analysis of the ψ backbone dihedral angle for one of the simulations are shown in the Appendix. The parameters obtained from the WLC fits are shown in Table (sim 2 and sim 3). The elastic properties obtained from these simulations at $v = 0.5\,\text{nm/ns}$ closely match those from the simulations at the slower pulling velocity of $v = 0.1\,\text{nm/ns}$. Thus, we consider the structural dynamics underlying the elastic properties of the polymers to be correctly captured already at the faster pulling velocity.

To test whether the system is sufficiently close to equilibrium during the simulation, we have performed both forward and backward pulling and analysed the hysteresis. As shown in the Appendix, there is virtually no hysteresis between the forward and backward force-extension traces, showing that the system was indeed close to equilibrium during the simulation.

To study finite size effects, we carried out force-probe MD simulations of a model system containig three instead of four azotripeptide units, (Lys-Azo-Gly)$_3$, at a pulling velocity of $0.1\,\text{nm/ns}$. Here, only all-*cis* and all-*trans* model polymers were simulated. The overall and individual force-extension curves are also shown in the Appendix. The elastic properties obtained from WLC fits to the force-extension traces of the shorter nonameric polymer are similar to those of the dodecamer ($L_\text{c} = 2.08\,\text{nm}$ and $L_\text{p} = 0.59\,\text{nm}$). Also in the nonamer, the lysine residues partly compensated the contraction of the azo groups upon optical *trans*-to-*cis* switching. Hence, we consider the dodecameric model polymer sufficiently as large and finite size effects as negligible.

7.3.5 Design of an Improved Polymer

The dependence of the extension of the lysine amino acids on the conformation of the azobenzene moiety to which the lysines are bonded to demonstrates that the peptides interlinking the photoactive azobenzenes are not just “passive glue”, but rather, through intermonomeric correlations, actively affect the elastic properties of the polymer. In the case at hand, the compensating effect of the lysines reduces the overall optical length contraction and thereby diminishes the work output of the polymer. We next investigated the possibility of using the structural and dynamical insight obtained from our studies to tailor and optimize photoswitchable polymers. For example, with the aim to increase the rigidity of the polymer backbone, we replaced the lysine and glycine residues with presumably stiffer proline residues, so that less compensation and, accordingly, larger work output could be expected.

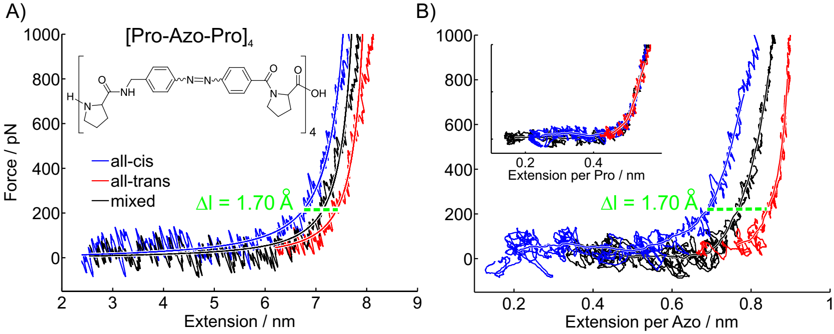


Figure 7.5: Force-probe MD force-extension curves for the (Pro-Azo-Pro)$_4$ polymer obtained at a pulling velocity of 0.5 nm/ns. **a)** Curves of the whole all-*cis* (blue), all-*trans* (red), and mixed *cis-trans-cis-trans* (black) polymers. Smooth lines describe the WLC fits. The extension difference between the *cis* and *trans* polymers is 1.70 Å/unit at 200 pN (dashed green line). **b)** Force-extension curves of individual azobenzene units and of proline residues (inset). The smooth lines represent 20,000-point averages.

Figure 7.5 shows the elastic behavior of the all-*cis*, all-*trans*, and mixed *cis-trans-cis-trans* (Pro-Azo-Pro)$_4$ dodecamers as determined from a set of force-probe MD simulations similar to those described above. The stiffer proline linkers indeed led to an increased extension difference between the *cis* and *trans* polymers of 1.70 Å per azo unit (dashed green line), as compared to 1.45 Å for the lysine-containing polymer (see Figure 7.3 and Table 7.3.4). Figure 7.5b shows the individual force-extension curves of the azobenzene and proline moieties in (Pro-Azo-Pro)$_4$. Apparently, and in contrast to (Lys-Azo-Gly)$_4$, the extension difference between the *cis* and *trans* azobenzene units is no longer compensated by the interlinking residues, which explains the increased Δl value of 1.70 Å. Accordingly, the work output is expected to be enhanced by about 15–20 % relative to that of the lysine-containing polymer.

7.4 Conclusions

The efficient and targeted optimization of photoswitchable polymers is a major challenge and requires a detailed microscopic interpretation of the polymer mechanics. We provided insight into the mechanics of photoswitchable polyazobenzene peptides under mechanical stress on the basis of force-probe molecular dynamics simulations. Comparison of the WLC parameters revealed that the overall elastic properties of the simulated azobenzene model polymers agree very well with the results of AFM experiments [89, 90]. Our simulations showed that the maximal obtainable extension difference between the *cis* and *trans* conformations is about 1.7 Å per switched azobenzene unit. However, in the AFM experiments, a value of only 0.6 Å was measured. By decomposing the force versus extension traces into the contributions of the individual building blocks, we identified the interlinking lysine amino acids as one reason for this discrepancy. The lysine linkers partly compensate the contraction and lengthening of the polymer upon optical *cis*-*trans* switching of the azobenzenes to 1.3–1.5 Å and thus are promising targets for optimizations. As an example, we constructed an optimized polyazobenzene peptide with stiffer proline linkers, which indeed led to a larger extension difference between the *cis* and *trans* forms and thus should enhance the work output by about 15–20 %.

Chapter 8

Summary and Conclusions

Understanding light-driven processes in condensed phase is a major goal of the bio- and nanosciences. The underlying molecular mechanisms in terms of the molecular dynamics are typically governed by sub-picosecond atomic motion. Mechanisms that occur at such ultra-fast timescales are very challenging to measure experimentally. The present thesis aimes at characterizing by means of molecular dynamics (MD) simulations the molecular mechanisms of three photochemical processes in condensed phase, the photoswitching mechanism of the fluoroprotein asFP595, the deactivation of an excited cytosine-guanine base pair in DNA, and the optical contraction of a photoswitchable polyazobenzene peptide. The simulations provided detailed structural and dynamic information about these processes at a resolution well beyond what is achievable experimentally. By using an *ab initio* QM/MM excited state MD strategy together with explicit surface hopping (asFP595 and DNA) or force-probe MD (polyazobenzene peptides), it was not only possible to quantitatively explain experimental results (such as quantum yields, excited state lifetimes, and force-extension curves), but also to make predictions that are rigorously testable, and in parts have already been tested, by experimental means. The detailed understanding of the molecular mechanism is a key step towards the rational improvement and design of photoactivatable systems, as exemplarily demonstrated for a polyazobenzene peptide.

Photoswitching Mechanism of the Fluoroprotein asFP595

The fluoroprotein asFP595 is a prototype of a reversibly switchable fluorescent protein. Its fluorescence can be switched *on* and *off* in response to irradiation of a particular wavelength. X-ray analysis showed that the *off-on* switching is accompanied by a conformational *trans-cis* isomerization of the chromophoric moiety within the protein matrix. However, the detailed mechanism underlying photoswitching was not understood at the atomistic level. In particular, it was unclear to what extent different protonation states of the chromophore and putative associated proton transfer events between the chromophore and the protein surrounding determine the optical properties of the protein.

The present thesis amined at elucidating the photoswitching mechanism of asFP595 (chapters 4 and 5). In chapter 4, in a first step, the influence of the protonation pattern of the chromophore and some proximal amino acids on the absorption characteristics of asFP595 was studied by means of combined force field MD simulations and quantum chemical as well as hybrid quantum/classical (QM/MM) calculations. In a second step (chapter 5), the photoinduced *trans-cis* photoisomerization dynamics of the chromophore within the protein matrix was simulated. To this end, QM/MM excited state MD simulations with explicit surface hopping at the conical intersection seam between the ground and excited states were carried out.

In the first step described in chapter 4, the simulated UV/Vis spectra were compared to the available experimental data (Figure 4.2). Together with computed protonation probabilities of the titratable groups in the chromophore cavity, the protonation patterns of the *off* and *on* states of asFP595 were unambiguously determined. In the *trans off* state, the zwitterionic and the anionic chromophores predominate, whereas the neutral chromophore form is populated in the *cis on* state. Thus, the simulations revealed that the *trans-cis* conformational switching of the chromophore is accompanied by proton transfer events. The different chromophore protonation in the *off* and the *on* state explains the experimentally observed absorption shift upon photoswitching.

An analysis of the hydrogen bonding network within the protein revealed that two distinct proton wires were formed that connect the chromophore cavity to the exterior solution (Figure 4.4). Along these proton wires, protons can enter to and exit from the chromophore cavity, thereby mediating the interconversion of the different chromophore protonation states during photoswitching. The wires involve several protonatable amino acid side chains and one buried crystallographic water molecule.

In the next step, the following key questions were addressed in chapter 5: How does light absorption induce the isomerization of the chromophore within the protein matrix, and how do the different protonation states affect the internal conversion mechanism? Which is the fluorescent species, and how can the fluorescence quantum yield be increased?

The simulations revealed that the different protonation patterns of the chromophore cavity govern the photoreactivity. Changes in the protonation state of the chromophore and some proximal amino acids lead to different photochemical states, which all are involved in the photoswitching process (Figure 5.9). These states are (i) the neutral chromophore species, which can undergo *trans-cis* photoisomerization, (ii) the anionic chromophores, which rapidly undergo radiationless decay after excitation, and (iii) the potentially fluorescent zwitterions. The overall stability of the different protonation states is controlled by the isomeric state of the chromophore.

Our excited state MD simulations showed that the zwitterionic chromophore is potentially fluorescent, irrespective of the conformation. However, we found an alternative deactivation channel for the *trans* zwitterion. This alternative channel does not involve isomerization, but rather excited state proton transfer from the zwitterion to the nearby glutamic acid Glu215. This deactivation pathway is not accessible for the *cis* chromophore. Only the latter is therefore trapped in the excited state and fluoresces. Furthermore, our simulations suggest that intense irradiation over a prolonged period of time could lead to a decarboxylation of the Glu215 side chain, thus rupturing the alternative channel. The absence of a radiationless deactivation channel leads to fluorescence also in the *trans* conformation, which provides a structural prediction of the irreversibly fluorescent state [7] of asFP595. Studying the proton transfers along the identified pathways in asFP595, both in the ground and the excited state, is the objective of future studies.

Taken together, the results presented in chapters 4 and 5 show that isomerization changes the preferred protonation state of the chromophore, which in turn determines the photochromic properties. These results suggest that a reversibly switchable protein must fulfill three criteria. First, to enable switching, *trans-cis* photoisomerization is necessary. Second, this photoisomerization has to be coupled to proton transfer events, that is, the preferred protonation state is different for the two conformers. Third, only one of the two conformers fluoresces while the other can undergo rapid radiationless decay. Recent experiments on Dronpa [302, 303] provide strong support for this proposed protonation/deprotonation mechanism. The similarity between the chromophores in a variety of fluoroproteins suggests that during molecular evolution, the (*p*-hydroxybenzylidene)imidazolinone chromophoric moiety served as a template and that the photochromic properties — and thus the function — were fine-tuned by the protein environment.

Deactivation of an Excited Cytosine-Guanine Base Pair in DNA

A second objective of the present thesis was to understand the origin of the intrinsic photostability of cytosine-guanine (C-G) Watson-Crick base pairs, both in the gas phase and embedded in DNA. This photostability results from a short excited state lifetime in the order of a few tens of femtoseconds [83] and might explain why DNA became the carrier of the genetic code as a result of selection pressure during molecular evolution.

In chapter 6, an extended conical intersection seam parallel to the proton transfer coordinate was identified (Figure 6.4) by means of excited state *ab initio* MD simulations and optimizations. This surface crossing seam explains the ultra-fast deactivation of the excited state, since decay can occur anywhere along the single proton transfer coordinate. In contrast to a previously proposed mechanism [85, 86], the proton transfer is not required for the decay, but rather a skeletal in-plane deformation of the base pair which drives the system towards the conical intersection seam.

The excited state lifetimes obtained from the simulations agree very well with the experiment. Furthermore, dynamic recrossings between the excited state and the ground state were observed in our simulations. Such recrossing phenomenon has also been observed experimentally for the diatomic sodium iodide gas [268, 269], which has only one internal degree of freedom. For large polyatomic systems such as the C-G base pair, however, such recrossings are an unusual phenomenon. An analysis of the minimum energy conical intersection revealed that the intersection has a rather unusual topology, in that there is only *one* degeneracy-lifting coordinate (instead of two), thus accounting for the observed partial diabatic trapping.

In the simulations, ultra-fast decay was observed both in the gas phase and in DNA. In the gas phase, however, decay occurred only after a complete proton transfer, whereas in DNA, decay can also occur much earlier. We found that the origin of this effect lies in the temporal electrostatic stabilization of the dipole in the charge transfer state in DNA.

By considering only a single base pair in the quantum mechanical calculations, we excluded the possibility of along-strand excimer formation. Recent experiments on an $(AT)_{18}$ B-DNA molecule have revealed that excited state relaxation mainly involves such intra-strand excimer formation [273]. Thus, also for DNA containing C-G base pairs, excimer states could play an important role in the decay process. Along-strand relaxation as well as the decay mechanism of the A-T base pair will be addressed in future investigations.

Elastic Properties of Photoswitchable Azobenzene Polymers

As the third contribution, the present thesis aimed at explaining the elastic properties of photoswitchable polyazobenzene peptides in terms of the underlying structural dynamics (chapter 7). Polyazobenzene peptides have been synthesized that can be optically contracted by means of light-induced *trans-cis* isomerization of the azobenzene units. In single-molecule atomic force microscopy (AFM) experiments, this contraction could even be established in an opto-mechanical operating cycle in which, in analogy to an Otto-cycle, repeated contraction-elongation against an external force delivered net mechanical work [89]. However, the work output of the polymer was considerably smaller than was expected on the basis of the length change of a single azobenzene monomer and the number of individual monomers. This reduced work output limits the applicability of the light-driven nanodevice as a molecular switch or cargo lifter.

We have carried out force-probe MD simulations of model polymers under mechanical stress that closely mimic the AFM experiments. Comparison of the wormlike chain parameters revealed that the overall elastic properties of the simulated model polymers agree very well with the results of the AFM experiments (Figure 7.3). Furthermore, the simulations explain the observed elastic properties in terms of the underlying structural dynamics at the atomistic scale. In particular, our force-probe simulations enabled to identify the lysine amino acids that interlink the azobenzene units as one reason for the reduced work output (Figure 7.4). In our simulations, the lysine linkers partly compensated the contraction and lengthening of the polymer upon optical *trans-cis* switching of the azobenzene units from 1.7 Å to 1.3–1.5 Å.

The insight obtained from the simulations is a prerequisite for the efficient and targeted optimization of photoswitchable polymers for future applications in nanotechnology. As a first example, we suggested a polyazobenzene peptide with stiffer proline linkers, which in our simulations indeed led to a larger extension difference between the *cis* and *trans* forms and thus should enhance the work output by about 15–20 %.

From a more general perspective, to ensure the proper functioning of a machine composed of force-generating units and interlinking units, the linkers have to be constructed in such a way as to sustain the mechanical stress generated. This thesis demonstrated in detail that this principle also holds for machines at the molecular level, such as photoswitchable azobenzene polymers.

Concluding Remarks

The investigation of three photoactivated processes in condensed phase by means of MD simulations, photoswitching of a fluoroprotein, ultrafast photodeactivation of a DNA base pair, and photo-induced changes of the elastic properties of polyazobenzene peptides, yielded a microscopic picture of the underlying mechanisms. By exploring the relevant regions of the free energy landscape at a high level of accuracy, our simulations shed light on the molecular properties of the chromophoric moiety and on the influence of the surrounding, which, taken together, guide the processes along the pathway. Thus, complemented by and cross-checked against experimental data, our simulations significantly contributed to the understanding of the respective function of the macromolecule in terms of the underlying structural dynamics.

To model the excited state dynamics correctly requires to fully characterize the temporal evolution from the instant of light absorption until deactivation. To achieve this goal, we used the complete active space self consistent field (CASSCF) quantum chemical method within a QM/MM scheme in conjunction with a diabatic surface hopping algorithm to describe the transitions between the excited and ground state energy surfaces. This methodological approach was developed recently by Gerrit Groenhof and co-workers, and was applied to study the photoactivation pathway in photoactive yellow protein [63]. From a methodological point of view, this thesis aimed at exploring how this new approach can be extended to fluorescent proteins and to nucleic acids. In particular, to study the fluoroprotein asFP595 required new methodological advances, namely the application of accelerated molecular dynamics simulation techniques, such as conformational flooding, which is detailed in chapters 2 and 3 of the present work.

Although the applications presented in this thesis clearly push the limits of what is feasible today for on-the-fly QM/MM molecular dynamics simulations, they also hint at current limitations. The calculation of CASSCF energies and forces severely constrains the size of the QM subsystem and enforces the use of minimal active spaces and basis sets. Therefore, prior to the MD simulations, the applied methods were calibrated against static *ab initio* calculations at higher levels of theory. For the applications described in the present work, the limited size of the QM subsystem impeded the direct simulation of possible excited state proton transfer processes in asFP595 and the study of along-strand relaxation pathways in DNA.

Thus, one main future challenge for theory, among others, is to develop efficient yet highly reliable methods for electronically excited states. In this respect, semi-empirical methods and valence-bond methods are very tempting due to their potential to yield results of similar quality as high-level methods at a drastically reduced computational cost. To achieve this goal, however, requires further methodological advances and careful parameterization. To enable an increasingly quantitative comparison of simulations with experiments, further advancements in the experimental assessment of the often ultra-fast molecular dynamics of photoactivated processes is desirable. Single-molecule techniques will undoubtedly play a key role in this endeavor.

Providing their comprehensive application and associated development, the potential of molecular dynamics simulations, as demonstrated in the present thesis, as a microscopic complement to experiments will then ultimately lead to a more fundamental understanding of photoactivated reactions at the atomistic level, and, furthermore, enable to predict and rationally modify the underlying processes.

Chapter 9

Appendix

9.1 asFP595

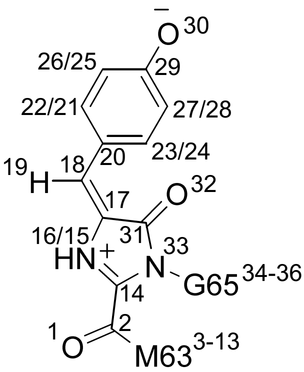


Figure 9.1: Schematic drawing of MYG defining the atom names used in the Tables.

atomic charges					
	type	atom	q_{trans,S_0}	q_{trans,S_1}	q_{cis,S_0}
1	opls_236	O	-0.419	-0.352	-0.369
2	opls_973	CA1	0.405	0.428	0.449
3	opls_136	CB1	-0.084	-0.086	-0.076
4	opls_140	HB1	0.060	0.060	0.060
5	opls_140	HB2	0.060	0.060	0.060
6	opls_210	CG1	0.098	0.098	0.098
7	opls_140	HG1	0.060	0.060	0.060
8	opls_140	HG2	0.060	0.060	0.060
9	opls_202	SD1	-0.435	-0.435	-0.435
10	opls_209	CE3	0.038	0.038	0.038
11	opls_140	HF1	0.060	0.060	0.060
12	opls_140	HF2	0.060	0.060	0.060
13	opls_140	HF3	0.060	0.060	0.060
14	opls_967	C1	0.217	0.144	0.111
15	opls_970	N2	-0.679	-0.484	-0.552
16	opls_513	H2	0.452	0.363	0.415
17	opls_968	CA2	0.354	0.171	0.266
18	opls_972	CB2	-0.504	-0.283	-0.365
19	opls_140	HB	0.179	0.144	0.134
20	opls_145	CG2	0.365	0.217	0.249
21	opls_145	CD1	-0.239	-0.185	-0.151
22	opls_146	HD1	0.142	0.133	0.160
23	opls_145	CD2	-0.229	-0.171	-0.184
24	opls_146	HD2	0.189	0.109	0.114
25	opls_145	CE1	-0.296	-0.251	-0.275
26	opls_146	HE1	0.146	0.123	0.125
27	opls_145	CE2	-0.303	-0.250	-0.257
28	opls_146	HE2	0.147	0.120	0.122
29	opls_166	CZ	0.786	0.671	0.694
30	opls_167	OH	-0.691	-0.618	-0.621
31	opls_969	C2	0.423	0.376	0.311
32	opls_236	O2	-0.546	-0.543	-0.525
33	opls_971	N3	-0.073	-0.014	-0.013
34	opls_223B	CA3	0.010	0.010	0.010
35	opls_140	HA1	0.064	0.054	0.054
36	opls_140	HA2	0.064	0.054	0.054

Table 9.2: OPLS parameters for MYG. Atomic charges were obtained from QM calculations of an isolated MYG model in the trans ground and excited states (q_{trans,S_0} and q_{trans,S_1}) and the cis ground state q_{cis,S_0}.

9.2 Azobenzene Polymers

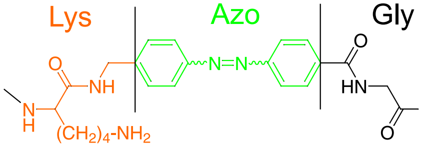


Figure 9.2: a) Schematic drawing of azotripeptide defining the decomposition into individual building blocks. The lysine, azobenzene, and glycine residues are shown in orange, green, and black, respectively. Note that the boundaries of the residues are different from the 'chemical boundaries' defined by the amide bonds, such that the azobenzene is solely defined by the N=N moiety plus the phenyl rings. **b)** Schematic drawing of azobenzene, defining the atom names used in the Tables below.

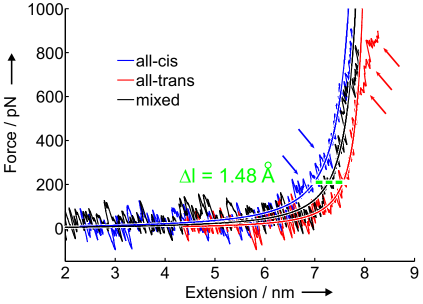


Figure 9.3: FPMD force-extension curves obtained at a pulling velocity of 0.5 nm/ns. The extension is defined as the distance between the C- and N-terminus of the polymer along the pulling direction. The curves for the all-cis, all-trans, and mixed polymers are shown in blue, red, and black, respectively. The data underlying the WLC fits (solid lines) were truncated at a force of 500 pN.

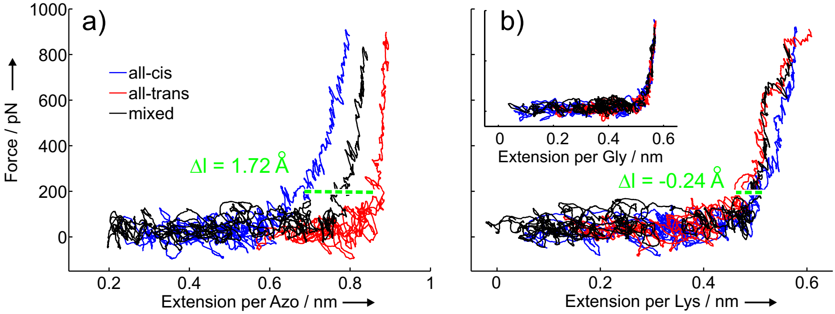


Figure 9.4: Force-extension curves of individual residues obtained at a pulling velocity of 0.5 nm/ns. (a) Average extension of azobenzenes. (b) Average extension of lysines and glycines (inset). The curves for the all-cis, all-trans, and mixed cis-trans-cis-trans dodecamers are shown in blue, red, and black, respectively.

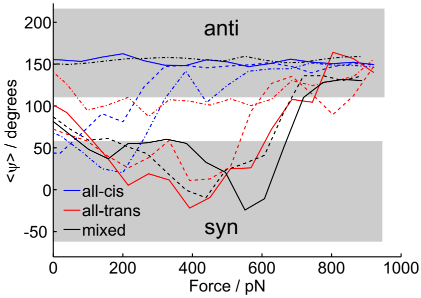


Figure 9.5: The average of the ψ dihedral angle is plotted as a function of the applied force for Lys4 (solid), Lys7 (dash-dotted), and Lys10 (dashed lines) of the all-cis (blue), all-trans (red), and mixed (black) polymers. The force was obtained at a pulling velocity of 0.5 nm/ns.

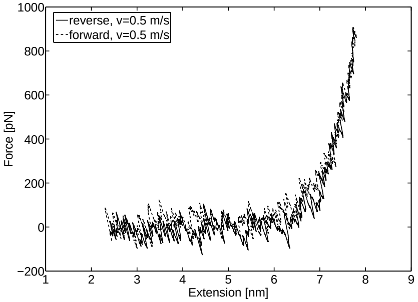


Figure 9.6: Forward (dashed line) and backward (solid line) FPMD force-extension curves of the all-cis polymer, obtained at a pulling velocity of 0.5 nm/ns. At this pulling velocity, there is virtually no hysteresis between the forward and backward directions, and thus the stretching is fully reversible.

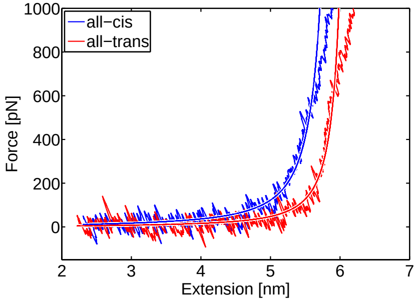


Figure 9.7: FPMD force-extension curves for the nonamer $(Lys - Azo - Gly)_3$, obtained at a pulling velocity of 0.1 nm/ns. The curves for the all-cis and all-trans systems are shown in blue and red, respectively.

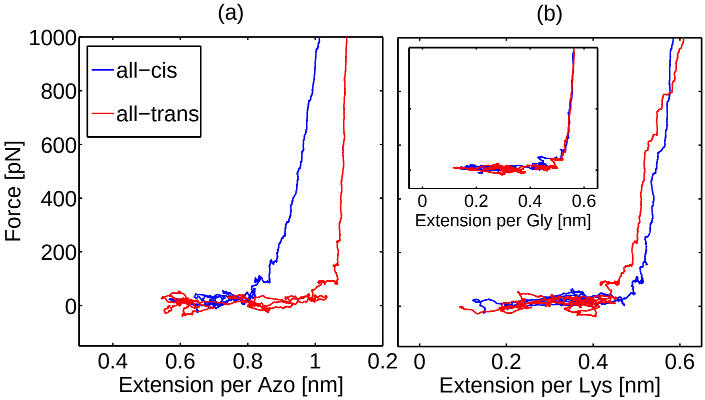


Figure 9.8: Force-extension curves of individual residues of the nonamer, obtained at a pulling velocity of 0.1 nm/ns. (a) Average extension of azobenzenes. (b) Average extension of lysines and glycines (inset). The curves for the all-cis and all-trans systems are shown in blue and red, respectively.

#	OPLS type	Atom type	q(cis)	q(trans)
6	opls_998 / 999	NAC / NAT	-0.28	-0.22
7	opls_145 / 997	CA / CAT	0.53	0.34
8	opls_145 / 997	CA / CAT	-0.33	-0.10
9	opls_145 / 997	CA / CAT	-0.02	-0.09
10	opls_145 / 997	CA / CAT	-0.14	-0.13
11	opls_145 / 997	CA / CAT	-0.02	-0.03
12	opls_145 / 997	CA / CAT	-0.33	-0.21
13	opls_998 / 999	NAC / NAT	-0.28	-0.22
14	opls_145 / 997	CA / CAT	0.53	0.34
15	opls_145 / 997	CA / CAT	-0.33	-0.10
16	opls_145 / 997	CA / CAT	-0.33	-0.21
17	opls_145 / 997	CA / CAT	-0.02	-0.09
18	opls_145 / 997	CA / CAT	-0.02	-0.03
19	opls_145 / 997	CA / CAT	0.04	0.07

Table 9.4: Atomic charges of azobenzene used for the force field simulations. Entries in the "OPLS type" and "Atom type" columns before and after the slash refer to cis and trans azobenzene, respectively. Standard OPLS types were used for atoms 1–5 and 20–29.

OPLS type	Atom type	V	W
opls_997	CAT	0.355	0.293076
opls_998	NAC	0.325	0.711756
opls_999	NAT	0.325	0.711756

Table 9.6: Added Lennard-Jones (6,12) parameters used in the OPLS force field of azobenzene.

Atom type	Atom type	b_0, nm	f_c, kJ/mol/nm^2
NAC	NAC	0.134	459710.6
NAT	NAT	0.118	459710.6
CA	NAC	0.143	357552.7
CA	NAT	0.140	357552.7

Table 9.8: Added equilibrium bond lengths and force constants used in the OPLS force field of azobenzene.

Atom type	Atom type	Atom type	θ_0, deg	f_c,kJ/mol/rad^2
CA	CA	NAC	124.0	586.152
CA	NAC	NAC	115.0	427.000
CAT	CAT	NAT	124.0	586.152
CAT	NAT	NAT	108.0	1000.000
CAT	CAT	CAT	120.0	700.000
CA	CA	CA	120.0	527.537
N	CT_2	CA	114.0	527.184
N	CT_2	CAT	114.0	527.184
CA	CT_2	N	114.0	527.184
CAT	CT_2	N	114.0	527.184
CT_2	CA	CA	120.0	585.760
CT_2	CAT	CAT	120.0	585.760
HC	CT_2	CA	109.5	292.880
HC	CT_2	CAT	109.5	292.880

Table 9.10: Equilibrium angles and force constants used in the OPLS force field of azobenzene.

type	type	type	type	C_0	C_1	C_2	C_3
NAC	NAC	CA	CA	16.038	-4.394	-14.394	2.395
CA	NAC	NAC	CA	100.000	0	-100.000	0
NAC	CA	CA	CA	30.354	0	-30.354	0
NAC	CA	CA	HA	30.354	0	-30.354	0
NAT	NAT	CAT	CAT	16.038	-4.394	-14.394	2.395
CAT	NAT	NAT	CAT	100.000	0	-100.000	0
NAT	CAT	CAT	CAT	30.354	0	-30.354	0
NAT	CAT	CAT	HA	30.354	0	-30.354	0
CAT	CAT	CAT	CAT	30.354	0	-30.354	0
CAT	CAT	CAT	HA	30.354	0	-30.354	0
HA	CAT	CAT	HA	30.354	0	-30.354	0
CAT	CAT	CAT	C	30.354	0	-30.354	0
HA	CAT	CAT	C	30.354	0	-30.354	0
CAT	CAT	C	O	8.782	0	-8.792	0
CAT	CAT	CAT	CT_2	30.354	0	-30.354	0
HA	CAT	CAT	CT_2	30.354	0	-30.354	0
CAT	CAT	C	N	4.605	0	-4.605	0

Table 9.12: Ryckaert-Bellemans dihedral parameters (given in kJ/mol) used in the OPLS force field of azobenzene. C_4 and C_5 were zero in all cases and thus are not listed.

Chapter 10

Acknowledgments

First of all, I am very grateful to Prof. Helmut Grubmüller for accepting me into his group at the MPI for Biophysical Chemistry and for his steady and great support during my time as a PhD student. He encouraged me to study asFP595 and azobenzene polymers, and he initiated fruitful collaborations with the groups of Prof. Stefan W. Hell (NanoBioPhotonics) and PD Markus C. Wahl (X-ray Crystallography) at the host institute, as well as with Prof. Hermann E. Gaub at the LMU München. During the evolution of the projects, he gave me the freedom to follow my own ideas, thereby giving me the opportunity to address questions which I find highly interesting.

I deeply thank Dr. Gerrit Groenhof for his support and for the good times in- and outside the lab. Gerrit supervised on a daily basis all QM/MM projects presented in this thesis. Without his masterly experience and advice, the present work would not have been possible.

I am grateful to Prof. Marcus Elstner for mentoring my thesis at the TU Braunschweig and for his advice, and Prof. Peter J. Walla for chairing my PhD committee.

I thank the Boehringer Ingelheim Foundation for a PhD scholarship and for the great meetings in Hirschegg and Blaubeuren, in particular Monika Beutelspacher, Claudia Walther, and Hermann Fröhlich. The B.I.F. also enabled stimulating visits of major conferences in Europe and overseas. Further financial support from the EU Nanomot project and the Deutsche Forschungsgemeinschaft is thankfully acknowledged.

I very much thank all former and current members of the Theoretical and Computational Biophysics department at the MPI for Biophysical

Chemistry for all the fruitful discussions and the advice, the highly motivating working environment, and the good fun. I especially thank Oliver Lange for his friendship and support, and Frauke Gräter for the collaboration during the initial phase of the asFP595 project. For reading parts of my thesis and for their invaluable comments, I thank Ira Tremmel and Guillem Portella. A very big and special thanks to Ira for helping me with the papers. Thanks to Martin Stumpe for his help with the cover, and to Matthias Müller for the support during my start-up phase in the group.

For the exceptionally fruitful collaboration and the support during my visits in London, I thank Dr. Martial Boggio-Pasqua and Prof. Mike Robb from the London Imperial College. Special thanks also to Dr. Astrid R. Klingen and Prof. Matthias Ullmann (Universität Bayreuth), Martin Andresen, Andre Stiel, Christian Eggeling, Markus Wahl, Stefan Jakobs, and Stefan Hell (MPI for Biophysical Chemistry) for the collaboration on asFP595.

A special thanks goes to Eveline Heinemann, Ansgar Esztermann, Martin Fechner, Ingo Hoffmann, and Oliver Slawik for adimistrative and technical support.

This thesis would not exist without the support from my family and friends. Thank you!

Bibliography

[1] L. Stryer. *Biochemistry.* W. H. Freeman and Co., New York, 3rd edition, 1988.

[2] D. Voet and J. G. Voet. *Biochemistry.* Wiley, New York, 2004.

[3] K. H. Kraemer. Sunlight and skin cancer: Another link revealed. *Proc. Natl. Acad. Sci. USA*, 94(1):11–14, 1997.

[4] R. P. Sinha and D. P. Hader. UV-induced DNA damage and repair: a review. *Photochem. Photobiol. Sci.*, 1(4):225–236, 2002.

[5] C. E. Crespo-Hernandez, B. Cohen, P. M. Hare and B. Kohler. Ultrafast excited-state dynamics in nucleic acids. *Chem. Rev.*, 104(4):1977–2019, 2004.

[6] R. Ando, H. Mizuno and A. Miyawaki. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science*, 306(5700):1370–1373, 2004.

[7] D. M. Chudakov, V. V. Belousov, A. G. Zarisky, V. V. Novoselov, D. B. Staroverov, D. B. Zorov, S. Lukyanov and K. A. Lukyanov. Kindling fluorescent proteins for precise in vivo photolabeling. *Nature Biotechnol.*, 21:191–194, 2003.

[8] K. A. Lukyanov, A. F. Fradkov, N. G. Gurskaya, M. V. Matz, Y. A. Labas, A. P. Savitsky, M. L. Markelov, A. G. Zaraisky, X. N. Zhao, Y. Fang, W. Y. Tan and S. A. Lukyanov. Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J. Biol. Chem.*, 275(34):25879–25882, 2000.

[9] T. A. Schüttrigkeit, T. von Feilitzsch, C. K. Kompa, K. A. Lukyanov, A. P. Savitsky, A. A. Voityuk and M. E. Michel-Beyerle.

Femtosecond study of light-induced fluorescence increase of the dark chromoprotein asFP595. *Chem. Phys.*, 323(2-3):149–160, 2006.

[10] J. M. Lehn. *Supramolecular Chemistry: Concepts and perspectives.* Wiley-VCH, Weinheim, 1995.

[11] J. P. Collin, P. Gavina, V. Heitz and J. P. Sauvage. Construction of one-dimensional multicomponent molecular arrays: Control of electronic and molecular motions. *Eur. J. Inorg. Chem.*, 1:1–14, 1998.

[12] V. Balzani, A. Credi, F. M. Raymo and J. F. Stoddart. Artificial molecular machines. *Angew. Chem.-Int. Edit.*, 39(19):3349–3391, 2000.

[13] B. L. Feringa, R. A. van Delden, N. Koumura and E. M. Geertsema. Chiroptical molecular switches. *Chem. Rev.*, 100(5):1789–1816, 2000.

[14] W. Hoppe, W. Lohmann, H. Markl and H. Ziegler, editors. *Biophysik.* Springer, 1982.

[15] J. W. Jung and W. Lee. Structure-based functional discovery of proteins: Structural proteomics. *J. Biochem. Mol. Biol.*, 37:28–34, 2004.

[16] A. T. Brunger and M. Nilges. Computational challenges for macromolecular structure determination by Xray crystallography and solution NMR spectroscopy. *Q. Rev. Biophys.*, 26:49–125, 1993.

[17] M. Nilges. Structure calculation from NMR data. *Curr. Opin. Struct. Biol.*, 6:617–623, 1996.

[18] E. Z. Eisenmesser, D. A. Bosco, M. Akke and D. Kern. Enzyme dynamics during catalysis. *Science*, 295(5559):1520–1523, 2002.

[19] R. Bruschweiler. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.*, 13(2):175–183, 2003.

[20] J. G. Kempf and J.P. Loria. Protein dynamics from solution NMR theory and applications. *Cell Biochem. Biophys.*, 37:187–211, 2003.

[21] J. C. Smith. Protein dynamics — comparison of simulations with inelastic neutron-scattering experiments. *Q. Rev. Biophys.*, 24:227–291, 1991.

[22] F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik and G. Zaccai. Protein dynamics studied by neutron scattering. *Q. Rev. Biophys.*, 35:327–367, 2002.

[23] H. J. Steinhoff. Methods for study of protein dynamics and protein-protein interaction in protein-ubiquitination by electron paramagnetic resonance spectroscopy. *Frontiers in Bioscience*, 7:C97–C110, 2002.

[24] H. Dietz and M. Rief. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. USA*, 103(5):1244–1247, 2006.

[25] H. Frauenfelder and P. G. Wolynes. Rate theories and puzzles of hemeprotein kinetics. *Science*, 229(4711):337–345, 1985.

[26] H. Frauenfelder, S. G. Sligar and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

[27] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA*, 102(19):6679–6685, 2005.

[28] S. Weiss. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.*, 7(9):724–729, 2000.

[29] S. T. Hess, S. H. Huang, A. A. Heikal and W. W. Webb. Biological and chemical applications of fluorescence correlation spectroscopy: A review. *Biochemistry*, 41(3):697–705, 2002.

[30] V. Srajer, T. Y. Teng, T. Ursby, C. Pradervand, Z. Ren, S. Adachi, W. Schildkamp, D. Bourgeois, M. Wulff and K. Moffat. Photolysis of the carbon monoxide complex of myoglobin: Nanosecond time-resolved crystallography. *Science*, 274(5293):1726–1729, 1996.

[31] V. Srajer, Z. Ren, T. Y. Teng, M. Schmidt, T. Ursby, D. Bourgeois, C. Pradervand, W. Schildkamp, M. Wulff and K. Moffat. Protein conformational relaxation and ligand migration in myoglobin: A nanosecond to millisecond molecular movie from time-resolved Laue X-ray diffraction. *Biochemistry*, 40(46):13802–13815, 2001.

[32] F. Garczarek and K. Gerwert. Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy. *Nature*, 439(7072):109–112, 2006.

[33] E. T. J. Nibbering, H. Fidder and E. Pines. Ultrafast chemistry: Using time-resolved vibrational spectroscopy for interrogation of structural dynamics. *Annu. Rev. Phys. Chem.*, 56:337–367, 2005.

[34] C. Kotting and K. Gerwert. Proteins in action monitored by time-resolved FTIR spectroscopy. *ChemPhysChem*, 6(5):881–888, 2005.

[35] J. Bredenbeck, J. Helbing, J. R. Kumita, G. A. Woolley and P. Hamm. Alpha-helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time-resolved IR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 102(7):2379–2384, 2005.

[36] H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim and A. Tokmakoff. Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin. *Proc. Natl. Acad. Sci. USA*, 102(3):612–617, 2005.

[37] Y. Xu, R. Oyola and F. Gai. Infrared study of the stability and folding kinetics of a 15-residue beta-hairpin. *J. Am. Chem. Soc.*, 125(50):15388–15394, 2003.

[38] R. Rammelsberg, G. Huhn, M. Lubben and K. Gerwert. Bacteriorhodopsin's intramolecular proton-release pathway consists of a hydrogen-bonded network. *Biochemistry*, 37(14):5001–5009, 1998.

[39] D. W. McCamant, P. Kukura and R. A. Mathies. Femtosecond stimulated Raman study of excited-state evolution in bacteriorhodopsin. *J. Phys. Chem. B*, 109(20):10449–10457, 2005.

[40] L. J. G. van Wilderen, W. van der Horst, M. A. van Stokkum, I. H. M. Hellingwerf, K. J. van Grondelle, R. Groot and M. L. van Wilderen. Ultrafast infrared spectroscopy reveals a key step for successful entry into the photocycle for photoactive yellow protein. *Proc. Natl. Acad. Sci. USA*, 103(41):15050–15055, 2006.

[41] A. H. Zewail. Femtochemistry. Past, present, and future. *Pure Appl. Chem.*, 72(12):2219–2231, 2000.

[42] S. K. Pal, J. Peon, B. Bagchi and A. H. Zewail. Biological water: Femtosecond dynamics of macromolecular hydration. *J. Phys. Chem. B*, 106(48):12376–12395, 2002.

[43] J. Peon, S. K. Pal and A. H. Zewail. Hydration at the surface of the protein monellin: Dynamics with femtosecond resolution. *Proc. Natl. Acad. Sci. USA*, 99(17):10964–10969, 2002.

[44] S. K. Pal and A. H. Zewail. Dynamics of water in biological recognition. *Chem. Rev.*, 104(4):2099–2123, 2004.

[45] A. Sharonov, T. Gustavsson, V. Carre, E. Renault and D. Markovitsi. Cytosine excited state dynamics studied by femtosecond fluorescence upconversion and transient absorption spectroscopy. *Chem. Phys. Lett.*, 380(1-2):173–180, 2003.

[46] D. Markovitsi, A. Sharonov, D. Onidas and T. Gustavsson. The effect of molecular organisation in DNA oligomers studied by femtosecond fluorescence spectroscopy. *ChemPhysChem*, 4(3):303, 2003.

[47] T. Gustavsson, A. Sharonov, D. Onidas and D. Markovitsi. Adenine, deoxyadenosine and deoxyadenosine 5'-monophosphate studied by femtosecond fluorescence upconversion spectroscopy. *Chem. Phys. Lett.*, 356(1-2):49–54, 2002.

[48] T. Gustavsson, A. Sharonov and D. Markovitsi. Thymine, thymidine and thymidine 5'-monophosphate studied by femtosecond fluorescence upconversion spectroscopy. *Chem. Phys. Lett.*, 351(3-4):195–200, 2002.

[49] J. Peon and A. H. Zewail. DNA/RNA nucleotides and nucleosides: direct measurement of excited-state lifetimes by femtosecond fluorescence up-conversion. *Chem. Phys. Lett.*, 348(3-4):255–262, 2001.

[50] J. M. L. Pecourt, J. Peon and B. Kohler. DNA excited-state dynamics: Ultrafast internal conversion and vibrational cooling in a series of nucleosides. *J. Am. Chem. Soc.*, 123(42):10370–10378, 2001.

[51] J. M. L. Pecourt, J. Peon and B. Kohler. Ultrafast internal conversion of electronically excited RNA and DNA nucleosides in water. *J. Am. Chem. Soc.*, 122(38):9348–9349, 2000.

[52] C. Desfrancois, S. Carles and J. P. Schermann. Weakly bound clusters of biological interest. *Chem. Rev.*, 100(11):3943–3962, 2000.

[53] A. Warshel and M. Levitt. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.

[54] Q. Cui, M. Elstner and M. Karplus. A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J. Phys. Chem. B*, 106:2721–2740, 2002.

[55] J. Gao and D. G. Truhlar. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.*, 53:467–505, 2002.

[56] A. Warshel. Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annu. Rev. Biophys. Struct.*, 32:425–443, 2003.

[57] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. B. Liu and M. H. M. Olsson. Electrostatic basis for enzyme catalysis. *Chem. Rev.*, 106(8):3210–3235, 2006.

[58] M. Casarin, D. Falcomer, A. Glisenti, M. M. Natile, F. Poli and A. Vittadini. Experimental and QM/MM investigation of the hydrated silica surface reactivity. *Chem. Phys. Lett.*, 405(4-6):459–464, 2005.

[59] S. A. French, A. A. Sokol, S. T. Bromley, C. R. A. Catlow and P. Sherwood. Identification and characterization of active sites and their catalytic processes - the Cu/ZnO methanol catalyst. *Top. Catal.*, 24(1-4):161–172, 2003.

[60] S. Namuangruk, P. Pantu and J. Limtrakul. Alkylation of benzene with ethylene over faujasite zeolite investigated by the ONIOM method. *J. Catal.*, 225(2):523–530, 2004.

[61] R. C. Deka and K. Hirao. Lewis acidity and basicity of cation-exchanged zeolites: QM/MM and density functional studies. *J. Mol. Catal. A-Chem.*, 181(1-2):275–282, 2002.

[62] A. H. deVries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto and G. J. Kramer. Zeolite structure and reactivity by combined quantum-chemical-classical calculations. *J. Phys. Chem. B*, 103(29):6133–6141, 1999.

[63] G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. de Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark and M. A. Robb.

Photoactivation of the Photoactive Yellow Protein: Why photon absorption triggers a trans-to-cis isomerization of the chromophore in the protein. *J. Am. Chem. Soc.*, 126:4228–4233, 2004.

[64] S. Hayashi, E. Tajkhorshid and K. Schulten. Molecular dynamics simulation of bacteriorhodopsin's photoisomerization using ab initio forces for the excited chromophore. *Biophys. J.*, 85(3):1440–1449, 2003.

[65] D. Marx. Proton transfer 200 years after von Grotthuss: Insights from ab initio simulations. *ChemPhysChem*, 7:1848–1870, 2006.

[66] W. Domcke, D. R. Yarkony and H. Köppel (eds.). *Conical Intersections, Advanced Series in Physical Chemistry Vol 15.* World Scientific, Singapore, 2004.

[67] K. Tai. Conformational sampling for the impatient. *Biophys. Chem.*, 107(3):213–220, 2004.

[68] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glaettli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt and H. B. Yu. Biomolecular modelling: goals, problems, perspectives. *Angewandte Chemie*, 45:4064–4092, 2006.

[69] H. Grubmüller. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E*, 52:2893–2906, 1995.

[70] E. M. Müller, A. de Meijere and H. Grubmüller. Predicting unimolecular chemical reactions: Chemical flooding. *J. Chem. Phys.*, 116(3):897–905, 2002.

[71] O. F. Lange, L. V. Schäfer and H. Grubmüller. Flooding in GROMACS: Accelerated barrier crossings in molecular dynamics. *J. Comp. Chem.*, 27:1693–1702, 2006.

[72] J. Lippincott-Schwartz, N. Altan-Bonnet and G. H. Patterson. Photobleaching and photoactivation: following protein dynamics in living cells. *Nat. Cell Biol.*, Pages S7–S14, 2003.

[73] A. Miyawaki, A. Sawano and T. Kogure. Lighting up cells: labelling proteins with fluorophores. *Nat. Cell Biol.*, Pages S1–S7, 2003.

[74] R. Y. Tsien. The green fluorescent protein. *Annu. Rev. Biochem.*, 67:509–544, 1998.

[75] J. C. Crano and R. J. Guglielmetti (eds.). *Organic Photochromic and Thermochromic Compounds: Main Photochromic Families, Vol. 1.* Plenum, New York, 1999.

[76] S. W. Hell, S. Jakobs and L. Kastrup. Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Appl. Phys. A-Mater. Sci. Process.*, 77(7):859–860, 2003.

[77] S. W. Hell. Toward fluorescence nanoscopy. *Nature Biotechnol.*, 21(11):1347–1355, 2003.

[78] S. W. Hell, M. Dyba and S. Jakobs. Concepts for nanoscale resolution in fluorescence microscopy. *Curr. Opin. Neurobiol.*, 14(5):599–609, 2004.

[79] M. Hofmann, C. Eggeling, S. Jakobs and S. W. Hell. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl. Acad. Sci. USA*, 102(49):17565–17569, 2005.

[80] M. Sauer. Reversible molecular photoswitches: A key technology for nanoscience and fluorescence imaging. *Proc. Natl. Acad. Sci. USA*, 102(27):9433–9434, 2005.

[81] J. D. Watson and F. H. C. Crick. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361):964–967, 1953.

[82] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids – a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[83] A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. deVries. Photochemical selectivity in guanine-cytosine base-pair structures. *Proc. Natl. Acad. Sci. USA*, 102(1):20–23, 2005.

[84] T. Schultz, E. Samoylova, W. Radloff, I. V. Hertel, A. L. Sobolewski and W. Domcke. Efficient deactivation of a model base pair via excited-state hydrogen transfer. *Science*, 306(5702):1765–1768, 2004.

[85] A. L. Sobolewski and W. Domcke. Ab initio studies on the photophysics of the guanine-cytosine base pair. *Phys. Chem. Chem. Phys.*, 6(10):2763–2771, 2004.

[86] A. L. Sobolewski, W. Domcke and C. Hättig. Tautomeric selectivity of the excited-state lifetime of guanine/cytosine base pairs: The role of electron-driven proton-transfer processes. *Proc. Natl. Acad. Sci. USA*, 102(50):17903–17906, 2005.

[87] G. S. Hartley. The cis-form of azobenzene. *Nature*, 140:281–281, 1937.

[88] H. Dürr and H. Bouas-Laurent (eds.). *H. Rau in: Photochromism: Molecules and Systems.* Elsevier, Amsterdam, 1990.

[89] T. Hugel, N. B. Holland, A. Cattani, L. Moroder, M. Seitz and H. E. Gaub. Single-molecule optomechanical cycle. *Science*, 296:1103–1106, 2002.

[90] N. B. Holland, T. Hugel, G. Neuert, A. Cattani-Scholz, C. Renner, D. Oesterhelt, L. Moroder, M. Seitz and H. E. Gaub. Single molecule force spectroscopy of azobenzene polymers: Switching elasticity of single photochromic macromolecules. *Macromolecules*, 36:2015–2023, 2003.

[91] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.*, 9(9):646–652, 2002.

[92] J. Gumbart, Y. Wang, A. Aksimentiev, E. Tajkhorshid and K. Schulten. Molecular dynamics simulations of proteins in lipid bilayers. *Curr. Opin. Struct. Biol.*, 15(4):423–431, 2005.

[93] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press, Oxford, 1989.

[94] D. Frenkel and B. Smit. *Understanding Molecular Simulation.* Academic Press, New York, 2002.

[95] F. Jensen. *Introduction to Computational Chemistry.* Wiley, New York, 2001.

[96] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry.* Dover Publications, New York, 1989.

[97] W. Koch and M. C. Holthausen. *A Chemist's Guide to Density Fuctional Theory.* Wiley-VCH, Weinheim, 2002.

[98] M. Born and R. Oppenheimer. Quantum theory of molecules. *Ann. d. Physik*, 84(20):457–484, 1927.

[99] R. W. Hockney. The potential calculation and some applications. *Methods Comput. Phys.*, 9:136–211, 1970.

[100] D. L. Beveridge J. A. Pople. *Approximate molecular orbital theory.* McGraw-Hill, New York, 1970.

[101] G. A. Segal. *Semiempirical Methods of electronic structure calculations.* Plenum Press, New York, 1977.

[102] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart. The development and use of quantum-mechanical molecular-models AM1 - a new general-purpose quantum-mechanical molecular-model. *J. Am. Chem. Soc.*, 107(13):3902–3909, 1985.

[103] W. Weber and W. Thiel. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.*, 103(6):495–506, 2000.

[104] T. Frauenheim, F. Weich, T. Kohler, S. Uhlmann, D. Porezag and G. Seifert. Density-functional-based construction of transferable nonorthogonal tight-binding potentials for Si and SiH. *Phys. Rev. B*, 52(15):11492–11501, 1995.

[105] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58(11):7260–7268, 1998.

[106] J. J. P. Stewart. Optimization of parameters for semiempirical methods. 1. Method. *J. Comput. Chem.*, 10(2):209–220, 1989.

[107] J. J. P. Stewart. Optimization of parameters for semiempirical methods. 2. Applications. *J. Comput. Chem.*, 10(2):221–264, 1989.

[108] W. L. Jorgensen and J. Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.

[109] J. M. Wang, P. Cieplak and P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21(12):1049–1074, 2000.

[110] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.

[111] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and user guide.* Vdf Hochschulverlag, Zürich, 1996.

[112] C. Oostenbrink, A. Villa, A. E. Mark and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53a5 and 53a6. *J. Comput. Chem.*, 25(13):1656–1676, 2004.

[113] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23(2):69–89, 1951.

[114] J. A. Pople, D. P. Santry and G. A. Segal. Approximate self-consistent molecular orbital theory. i. Invariant procedures. *J. Chem. Phys.*, 43(10P2):S129, 1965.

[115] J. A. Pople and G. A. Segal. Approximate self-consistent molecular orbital theory. 2. Calculations with complete neglect of differential overlap. *J. Chem. Phys.*, 43(10P2):S136, 1965.

[116] J. A. Pople, P. A. Beveridge and P. A. Dobosh. Approximate self-consistent molecular-orbital theory. 5. Intermediate neglect of differential overlap. *J. Chem. Phys.*, 47(6):2026, 1967.

[117] R. Pariser and R. G. Parr. A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. II. *J. Chem. Phys.*, 21(5):767–776, 1953.

[118] R. Pariser and R. G. Parr. A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. I. *J. Chem. Phys.*, 21(3):466–471, 1953.

[119] J. Ridley and M. C. Zerner. Intermediate neglect of differential overlap technique for spectroscopy - pyrrole and azines. *Theor. Chim. Acta*, 32(2):111–134, 1973.

[120] L. K. Hanson, J. Fajer, M. A. Thompson and M. C. Zerner. Electrochromic effects of charge separation in bacterial photosynthesis - theoretical-models. *J. Am. Chem. Soc.*, 109(15):4728–4730, 1987.

[121] W. P. Anderson, W. D. Edwards and M. C. Zerner. Calculated spectra of hydrated ions of the 1st transition-metal series. *Inorg. Chem.*, 25(16):2728–2732, 1986.

[122] G. L. Estiu and M. C. Zerner. Calculations on the electronic structure and UV-visible spectrum of oxyhemocyanin. *J. Am. Chem. Soc.*, 121(9):1893–1901, 1999.

[123] C. J. Cramer. *Essentials of Computational Chemistry.* Wiley, New York, 2002.

[124] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev. B*, 136(3B):B864, 1964.

[125] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):1133, 1965.

[126] E. Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52(12):997–1000, 1984.

[127] C. Jamorski, M. E. Casida and D. R. Salahub. Dynamic polarizabilities and excitation spectra from a molecular implementation of time-dependent density-functional response theory: N-2 as a case study. *J. Chem. Phys.*, 104(13):5134–5147, 1996.

[128] M. Petersilka, U. J. Gossmann and E. K. U. Gross. Excitation energies from time-dependent density-functional theory. *Phys. Rev. Lett.*, 76(8):1212–1215, 1996.

[129] A. Dreuw and M. Head-Gordon. Failure of time-dependent density functional theory for long-range charge-transfer excited states:

The zincbacteriochlorin-bacterlochlorin and bacteriochlorophyll-spheroidene complexes. *J. Am. Chem. Soc.*, 126(12):4007–4016, 2004.

[130] C. P. Hsu, S. Hirata and M. Head-Gordon. Excitation energies from time-dependent density functional theory for linear polyene oligomers: Butadiene to decapentaene. *J. Phys. Chem. A*, 105(2):451–458, 2001.

[131] P. Siegbahn, A. Heiberg, B. Roos and B. Levy. Comparison of the super-CL and the Newton-Raphson scheme in the complete active space SCF method. *Phys. Scr.*, 21(3-4):323–327, 1980.

[132] P. E. M. Siegbahn, J. Almlof, A. Heiberg and B. O. Roos. The complete active space SCF (CASSCF) method in a Newton-Raphson formulation with application to the HNO molecule. *J. Chem. Phys.*, 74(4):2384–2396, 1981.

[133] J. J. W. McDouall, K. Peasley and M. A. Robb. A simple MCSCF perturbation theory – orthogonal valence bond Moller-Plesset-2 (OVB-MP2). *Chem. Phys. Lett.*, 148(2-3):183–189, 1988.

[134] K. Andersson, P. A. Malmqvist and B. O. Roos. 2nd-order perturbation-theory with a complete active space self-consistent field reference function. *J. Chem. Phys.*, 96(2):1218–1226, 1992.

[135] M. A. Robb, F. Bernardi and M. Olivucci. Conical intersections as a mechanistic feature of organic-photochemistry. *Pure Appl. Chem.*, 67(5):783–789, 1995.

[136] F. Bernardi, M. Olivucci and M. A. Robb. Potential energy surface crossings in organic photochemistry. *Chem. Soc. Rev.*, 25(5):321, 1996.

[137] M. E. Martin, F. Negri and M. Olivucci. Origin, nature, and fate of the fluorescent state of the green fluorescent protein chromophore at the CASPT2//CASSCF resolution. *J. Am. Chem. Soc.*, 126:5452–5264, 2004.

[138] G. Groenhof. *Understanding light-induced conformational changes in molecular systems from first principles.* University of Groningen, Groningen, 2005.

[139] G. Herzberg and H. C. Longuet-Higgins. Intersection of potential energy surfaces in polyatomic molecules. *Discus. Farad. Soc.*, 35:77, 1963.

[140] H. C. Longuet-Higgins. Intersection of potential-energy surfaces in polyatomic-molecules. *Proc. R. Soc. London Ser. A-Math. Phys. Eng. Sci.*, 344(1637):147–156, 1975.

[141] K. B. Lipkowitz and D. B. Boyd (eds) M. A. Robb and M. Garavelli and M. Olivucci and F. Bernardi in. *Reviews in Computational Chemistry.* Wiley, New York, 2000.

[142] G. A. Worth and M. A. Robb. Applying direct molecular dynamics to non-adiabatic systems. *Adv. Chem. Phys.*, 124:355–431, 2002.

[143] G. A. Worth and L. S. Cederbaum. Beyond Born-Oppenheimer: Molecular dynamics through a conical intersection. *Annu. Rev. Phys. Chem.*, 55:127–158, 2004.

[144] K. Gerwert, G. Souvignier and B. Hess. Simultaneous monitoring of light-induced-changes in protein side-group protonation, chromophore isomerization, and backbone motion of bacteriorhodopsin by time-resolved Fourier-transform infrared-spectroscopy. *Proc. Natl. Acad. Sci. USA*, 87(24):9774–9778, 1990.

[145] A. Warshel and Z. T. Chu. Nature of the surface crossing process in bacteriorhodopsin: Computer simulations of the quantum dynamics of the primary photochemical event. *J. Phys. Chem. B*, 105:9857–9871, 2001.

[146] A. N. Bondar, M. Elstner, S. Suhai, J. C. Smith and S. Fischer. Mechanism of primary proton transfer in bacteriorhodopsin. *Structure*, 12(7):1281–1288, 2004.

[147] N. Agmon. Proton pathways in green fluorescence protein. *Biophys. J.*, 88(4):2452–2461, 2005.

[148] D. Stoner-Ma, A. A. Jaye, P. Matousek, M. Towrie, S. R. Meech and P. J. Tonge. Observation of excited-state proton transfer in green fluorescent protein using ultrafast vibrational spectroscopy. *J. Am. Chem. Soc.*, 127(9):2864–2865, 2005.

[149] V. Helms. Electronic excitations of biomolecules studied by quantum chemistry. *Curr. Opin. Struct. Biol.*, 12(2):169–175, 2002.

[150] J. von Neumann and E. Wigner. Concerning the behaviour of eigenvalues in adiabatic processes. *Z. Physik*, 30:467–470, 1929.

[151] H. Köppel, W. Domcke and L. S. Cederbaum. Multimode molecular-dynamics beyond the born-oppenheimer approximation. *Advances in Chemical Physics*, 57:59–246, 1984.

[152] M. Thachuk, M. Y. Ivanov and D. M. Wardlaw. A semiclassical approach to intense-field above-threshold dissociation in the long wavelength limit. II. Conservation principles and coherence in surface hopping. *J. Chem. Phys.*, 109(14):5747–5760, 1998.

[153] D. Bakowies and W. Thiel. Semiempirical treatment of electrostatic potentials and partial charges in combined QM/MM approaches. *J. Comp. Chem.*, 17(1):87–108, 1996.

[154] F. Maseras and K. Morokuma. IMOMM - a new integrated abinitio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition-states. *J. Comput. Chem.*, 16(9):1170–1179, 1995.

[155] J. Sauer and M. Sierka. Combining quantum mechanics and interatomic potential functions in ab initio studies of extended systems. *J. Comput. Chem.*, 21(16):1470–1493, 2000.

[156] P. Amara and M. J. Field. Evaluation of an ab initio quantum mechanical / molecular mechanical hybrid-potential link-atom method. *Theor. Chem. Acc.*, 109:43–52, 2003.

[157] M. A. Thompson and G. K. Schenter. Excited-states of the bacteriochlorophyll-b dimer of rhodopseudomonas-viridis - a QM/MM study of the photosynthetic reaction-center that includes MM polarization. *J. Phys. Chem.*, 99(17):6374–6386, 1995.

[158] A. Mallik, D. E. Taylor, K. Runge, J. W. Dufty and H. P. Cheng. Procedure for building a consistent embedding at the QM-CM interface. *J. Comput-Aided Mater. Des.*, 13(1-3):45–60, 2006.

[159] C. J. R. Illingworth, S. R. Gooding, P. J. Winn, G. A. Jones, G. G. Ferenczy and C. A. Reynolds. Classical polarization in hybrid QM/MM methods. *J. Phys. Chem. A*, 110:6487–6497, 2006.

[160] M. J. Field, P. A. Bash and M. Karplus. A combined QM/MM potential for MD simulations. *J. Comp. Chem.*, 11(6):700–733, 1990.

[161] U. C. Singh and P. A. Kollman. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl- exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, 7(6):718–730, 1986.

[162] G. G. Ferenczy, J. L. Rivail, P. R. Surjan and G. Naray-Szabo. NDDO fragment self-consistent field approximation for large electronic systems. *J. Comput. Chem.*, 13(7):830–837, 1992.

[163] D. M. Philipp and R. A. Friesner. Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *J. Comput. Chem.*, 20(14):1468–1494, 1999.

[164] J. Gao, P. Amara, C. Alhambra and M. J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem A*, 102:4714–4721, 1998.

[165] H. Grubmüller, B. Heymann and P. Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999, 1996.

[166] B. Isralewitz, S. Izrailev and K. Schulten. Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations. *Biophys. J.*, 73(6):2972–2979, 1997.

[167] U. Nienhaus (Ed) H. Grubmüller in. *Protein–Ligand Interactions.* The Humana Press Inc., Totowa, NJ, 2005.

[168] B. Heymann and H. Grubmüller. Chair-boat transitions and side groups affect the stiffness of polysaccharides. *Chem. Phys. Lett.*, 305(3-4):202–208, 1999.

[169] F. Gräter, J. Shen, H. Jiang, M. Gautel and H. Grubmüller. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys. J.*, 88:790–804, 2005.

[170] H. B. Schlegel. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *J. Comput. Chem.*, 24(12):1514–1527, 2003.

[171] T. Huber, A. E. Torda and W. F. van Gunsteren. Local elevation - a method for improving the searching properties of molecular-dynamics simulation. *J. Comput.-Aided Mol. Des.*, 8(6):695–708, 1994.

[172] A. F. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.*, 78(20):3908–3911, 1997.

[173] A. F. Voter. A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys.*, 106(11):4665–4677, 1997.

[174] D. Hamelberg, J. Mongan and J. A. McCammon. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, 120(24):11919–11929, 2004.

[175] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99(20):12562–12566, 2002.

[176] A. Kitao, F. Hirata and N. Gō. The effects of solvent on the conformation and the collective motions of protein - normal mode analysis and molecular-dynamics simulations of melittin in water and in vacuum. *Chem. Phys.*, 158(2-3):447–472, 1991.

[177] A. E. Garcia. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 68(17):2696–2699, 1992.

[178] A. Amadei, A. B. M. Linssen and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins*, 17(4):412–425, 1993.

[179] M. Levitt, C. Sander and P. S. Stern. The normal-modes of a protein - native bovine pancreatic trypsin-inhibitor. *Int. J. Quantum Chem.*, Suppl. 10:181–199, 1983.

[180] B. Brooks and M. Karplus. Harmonic dynamics of proteins - normal-modes and fluctuations in bovine pancreatic trypsin-inhibitor. *Proc. Natl. Acad. Sci. USA*, 80(21):6571–6575, 1983.

[181] N. Gō, T. Noguti and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational-modes. *Proc. Natl. Acad. Sci. USA*, 80(12):3696–3700, 1983.

[182] A. Amadei, B. L. de Groot, M. A. Ceruso, M. Paci, A. Di Nola and H. J. C. Berendsen. A kinetic model for the internal motions

of proteins: Diffusion between multiple harmonic wells. *Proteins: Struct. Funct. Genet.*, 35:283–292, 1999.

[183] D. van der Spoel, B. Hess, E. Lindahl, G. Groenhof, A. E. Mark and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J. Comput. Chem.*, 26(16):1701–1718, 2005.

[184] B. L. de Groot, G. Vriend and H. J. C. Berendsen. Conformational changes in the chaperonin GroEL: New insights into the allosteric mechanism. *J. Mol. Biol.*, 286(4):1241–1249, 1999.

[185] M. J. Frisch *et al.* Gaussian 03, Revision B.04. Gaussian, Inc., Wallingford, CT, 2004.

[186] K. P. Vollhardt and N. E. Schore. *Organic Chemistry.* W H Freeman, New York, 1998.

[187] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[188] S. Miyamoto and P. A. Kollman. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.*, 13:952–962, 1992.

[189] M. Nendel, D. Sperling, O. Wiest and K. N. Houk. Computational explorations of vinylcyclopropane-cyclopentene rearrangements and competing diradical stereoisomerizations. *J. Org. Chem.*, 65(11):3259–3268, 2000.

[190] B. Engels, M. Hanrath and C. Lennartz. Individually selecting multi-reference CI and its application to biradicalic cyclizations. *Computers and Chemistry*, 25(1):15–38, 2001.

[191] E. Nakamura, S. Yamago, S. Ejiri, A. E. Dorigo and K. Morokuma. Reversible generation of trimethylenemethanes by mild thermolysis of dialkoxymethylenecyclopropanes. *J. Am. Chem. Soc.*, 113:3183–3184, 1991.

[192] O. F. Lange and H. Grubmüller. Can principal components yield a dimension reduced description of dynamics on long time scales? *J. Phys. Chem. B*, 110:22842–22852, 2006.

[193] B. Hess. Convergence of sampling in protein simulations. *Phys. Rev. E*, 65(3):031910, 2002.

[194] J. D. Faraldo-Gomez, L. R. Forrest, M. Baaden, P. J. Bond, C. Domene, G. Patargias, J. Cuthbertson and M. S. P. Sansom. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins*, 57(4):783–791, 2004.

[195] B. Bouvier and H. Grubmüller. Molecular dynamics study of slow base flipping in DNA using conformational flooding. *submitted*, 2007.

[196] M. A. Balsera, W. Wriggers, Y. Oono and K. Schulten. Principal component analysis and long time protein dynamics. *J. Phys. Chem.*, 100(7):2567–2572, 1996.

[197] A. Salih, A. Larkum, G. Cox, M. Kühl and O. Hoegh-Guldberg. Fluorescent pigments in corals are photoprotective. *Nature*, 408:850–853, 2000.

[198] D. Veith and M. Veith. Biologie fluoreszierender Proteine. *Biol. Unserer Zeit*, 6(35):394–404, 2005.

[199] S. Habuchi, R. Ando, P. Dedecker, W. Verheijen, H. Mizuno, A. Miyakawa and J. Hofkens. Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proc. Natl. Acad. Sci. USA*, 102(27):9511–9516, 2005.

[200] P. G. Wilmann, K. Turcic, J. M. Battad, M. C. J. Wilce, R. J. Devenish, M. Prescott and J. Rossjohn. The 1.7 Å crystal structure of Dronpa: A photoswitchable green fluorescent protein. *J. Mol. Biol.*, 364(2):213–224, 2006.

[201] P. Dedecker, J. Hotta, R. Ando, A. Miyakawa, Y. Engelborghs and J. Hofkens. Fast and reversible photoswitching of the fluorescent protein Dronpa as evidenced by fluorescence correlation spectroscopy. *Biophys. J.*, 91(5):L45–L47, 2006.

[202] S. J. Remington. Fluorescent proteins: maturation, photochemistry and photophysics. *Curr. Opin. Struct. Biol.*, 16(6):714–721, 2006.

[203] K. A. Lukyanov, D. M. Chudakov, S. Lukyanov and V. V. Verkhusha. Photoactivatable fluorescent proteins. *Nat. Rev. Mol. Cell Biol.*, 6(11):885–891, 2005.

[204] D. M. Chudakov, A. V. Feofanov, N. N. Mudrik, S. Lukyanov and K. A. Lukyanov. Chromophore environment provides clue to "kindling fluorescent protein" riddle. *J. Biol. Chem.*, 278(9):7215–7219, 2003.

[205] M. Andresen, M. C. Wahl, A. C. Stiel, F. Gräter, L. V. Schäfer, S. Trowitzsch, G. Weber, C. Eggeling, H. Grubmüller, S. W. Hell and S. Jakobs. Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 102(37):13070–13074, 2005.

[206] P. G. Wilmann, J. Petersen, R. J. Devenish, M. Prescott and J. Rossjohn. Variations on the GFP chromophore. *J. Biol. Chem.*, 280(4):2401–2404, 2005.

[207] M. L. Quillin, D. A. Anstrom, X. K. Shu, S. O'Leary, K. Kallio, D. A. Chudakov and S. J. Remington. Kindling fluorescent protein from anemonia sulcata: Dark-state structure at 1.38 Å resolution. *Biochemistry*, 44(15):5774–5787, 2005.

[208] I. V. Yampolsky, S. J. Remington, V. I. Martynov, V. K. Potapov, S. Lukyanov and K. A. Lukyanov. Synthesis and properties of the chromophore of the asFP595 chromoprotein from Anemonia sulcata. *Biochemistry*, 44(15):5788–5793, 2005.

[209] P. Leiderman, D. Huppert and N. Agmon. Transition in the temperature-dependence of GFP fluorescence: From proton wires to proton exit. *Biophys. J.*, 90(3):1009–1018, 2006.

[210] O. Vendrell, R. Gelabert, M. Moreno and J. M. Lluch. Potential energy landscape of the photoinduced multiple proton-transfer process in the green fluorescent protein: classical molecular dynamics and multiconfigurational electronic structure calculations. *J. Am. Chem. Soc.*, 128:3564–3574, 2006.

[211] A. D. Bacon and M. C. Zerner. Intermediate neglect of differential overlap theory for transition-metal complexes - Fe, Co and Cu chlorides. *Theor. Chim. Acta*, 53(1):21–54, 1979.

[212] M. E. Casida and D. P. Chong, ed. *Recent Advances in Density Functional Methods.* World Scientific Publishing Co. Ltd., Singapore, 1995.

[213] L. N. Oliveira, E. K. U. Gross and W. Kohn. Ensemble-density functional theory for excited-states. *Int. J. Quant. Chem.*, Suppl. 24:707–716, 1990.

[214] G. Groenhof, M. F. Lensink, H. J. C. Berendsen, J. G. Snijders and A. E. Mark. Signal transduction in the photoactive yellow protein. I. Photon absorption and the isomerization of the chromophore. *Proteins*, 48(2):202–211, 2002.

[215] M. A. L. Marques, X. Lopez, D. Varsano, A. Castro and A. Rubio. Time-dependent density-functional approach for biological chromophores: The case of the green fluorescent protein. *Phys. Rev. Lett.*, 90(25), 2003.

[216] J. A. Gascon and V. S. Batista. QM/MM study of energy storage and molecular rearrangements due to the primary event in vision. *Biophys. J.*, 87(5):2931–2941, 2004.

[217] X. Lopez, M. A. L. Marques, A. Castro and A. Rubio. Optical absorption of the blue fluorescent protein: A first-principles study. *J. Am. Chem. Soc.*, 127(35):12329–12337, 2005.

[218] N. Sanna, G. Chillemi, A. Grandi, S. Castelli, A. Desideri and Barone. New hints on the pH-driven tautomeric equilibria of the topotecan anticancer drug in aqueous solutions from an integrated spectroscopic and quantum-mechanical approach. *J. Am. Chem. Soc.*, 127(44):15429–15436, 2005.

[219] B. Grigorenko, A. Savitsky, I. Topol, S. Burt and A. Nemukhin. trans and cis chromophore structures in the kindling fluorescent protein asFP595. *Chem. Phys. Lett.*, 424:184–188, 2006.

[220] S. H. Vosko, L. Wilk and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin-density calculations - a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, 1980.

[221] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A*, 38(6):3098–3100, 1988.

[222] J. P. Perdew. Density-functional approximation for the correlation-energy of the inhomogeneous electron-gas. *Phys. Rev. B*, 33(12):8822–8824, 1986.

[223] A. F. Bell, X. He, R. M. Wachter and P. J. Tonge. Probing the ground state structure of the Green Fluorescent Protein chromophore using Raman spectroscopy. *Biochemistry*, 39(15):4423–4431, 2000.

[224] R. W. W. Hooft, C. Sander and G. Vriend. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins — Structure, Function, and Genetics*, 26(4):363–376, 1996.

[225] G. Vriend. WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.

[226] B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.

[227] T. Darden, D. York and L. Pedersen. Particle Mesh Ewald — An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[228] C. M. Breneman and K. B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials – the need for high sampling density in formamide conformational-analysis. *J. Comput. Chem.*, 11(3):361–373, 1990.

[229] S. J. A. van Gisbergen, F. Kootstra, P. R. T. Schipper, O. V. Gritsenko, J. G. Snijders and E. J. Baerends. Density-functional-theory response-property calculations with accurate exchange-correlation potentials. *Phys. Rev. A*, 57(4):2556–2571, 1998.

[230] S. J. A. van Gisbergen, A. Rosa, G. Ricciardi and E. J. Baerends. Time-dependent density functional calculations on the electronic absorption spectrum of free base porphin. *J. Chem. Phys.*, 111(6):2499–2506, 1999.

[231] A. V. Nemukhin, I. A. Topol and S. K. Burt. Electronic excitations of the chromophore from the fluorescent protein asFP595 in solutions. *J. Chem. Theory Comput.*, 2:292–299, 2006.

[232] G. M. Ullmann and E. W. Knapp. Electrostatic models for computing protonation and redox equilibria in proteins. *Eur. Biophys. J.*, 28:533–551, 1999.

[233] D. Bashford and M. Karplus. pKas of ionizable groups in proteins - atomic detail from a continuum electrostatic model. *Biochemistry*, 29(44):10219–10225, 1990.

[234] G. M. Ullmann. *QMCT: a Monte Carlo titration program*, 2005.

[235] A. Bondi. Van der Waals volumes + radii. *J. Phys. Chem.*, 68(3):441, 1964.

[236] D. Bashford, D. A. Case, C. Dalvit, L. Tennent and P. E. Wright. Electrostatic calculations of side-chain pka values in myoglobin and comparison with NMR data for histidines. *Biochemistry*, 32:8045–8056, 1993.

[237] M. D. Liptak and G. C. Shields. Accurate pKa calculations for carboxylic acids using complete basis set and gaussian-n models combined with CPCM continuum solvation methods. *J. Am. Chem. Soc.*, 123:7314–7319, 2001.

[238] S. Miertus, E. Scrocco and J. Tomasi. Electrostatic interaction of a solute with a continuum - a direct utilization of abinitio molecular potentials for the prevision of solvent effects. *Chem. Phys.*, 55(1):117–129, 1981.

[239] M. Cossi, V. Barone, R. Cammi and J. Tomasi. Ab initio study of solvated molecules: A new implementation of the polarizable continuum model. *Chem. Phys. Lett.*, 255(4-6):327–335, 1996.

[240] E. Cances, B. Mennucci and J. Tomasi. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.*, 107(8):3032–3041, 1997.

[241] V. Barone, M. Cossi and J. Tomasi. Geometry optimization of molecular structures in solution by the polarizable continuum model. *J. Comput. Chem.*, 19(4):404–417, 1998.

[242] M. Cossi, V. Barone, B. Mennucci and J. Tomasi. Ab initio study of ionic solutions by a polarizable continuum dielectric model. *Chem. Phys. Lett.*, 286(3-4):253–260, 1998.

[243] A. R. Klingen and G. M. Ullmann. Negatively charged residues and hydrogen bonds tune the ligand histidine pka values of Rieske iron-sulfur proteins. *Biochemistry*, 43(39):12383–12389, 2004.

[244] M. Wanko, M. Garavelli, F. Bernardi, T. A. Niehaus, T. Frauenheim and M. Elstner. A global investigation of excited state surfaces within time-dependent density-functional response theory. *J. Chem. Phys.*, 120(4):1674–1692, 2004.

[245] T. Laino, R. Nifosi and V. Tozzini. Relationship between structure and optical properties in green fluorescent proteins: a quantum mechanical study of the chromophore environment. *Chem. Phys.*, 298(1-3):17–28, 2004.

[246] S. Habuchi, M. Cotlet, T. Gensch, T. Bednarz, S. Haber-Pohlmeier, J. Rozenski, G. Dirix, J. Michiels, J. Vanderleyden, J. Heberle, F. C. DeSchryver and J. Hofkens. Evidence for the isomerization and decarboxylation in the photoconversion of the red fluorescent protein DsRed. *J. Am. Chem. Soc.*, 127(25):8977–8984, 2005.

[247] G. U. Nienhaus, K. Nienhaus, A. Holzle, S. Ivanchenko, F. Renzi, F. Oswald, M. Wolff, F. Schmitt, C. Rocker, B. Vallone, W. Weidemann, R. Heilker, H. Nar and J. Wiedenmann. Photoconvertible fluorescent protein EosFP: Biophysical properties and cell biology applications. *Photochem. Photobiol.*, 82(2):351–358, 2006.

[248] B. Bieger, L.-O. Essen and D. Oesterhelt. Crystal structure of halophilic dodecin: A novel, dodecameric flavin binding protein from halobacterium salinarum. *Structure*, 11:375–385, 2003.

[249] N. M. Derix, R. W. Wechselberger, M. A. van der Horst, K. J. Hellingwerf, R. Boelens, R. Kaptein and N. A. J. van Nuland. Lack of negative charge in the E46Q mutant of photoactive yellow protein prevents partial unfolding of the blue-shifted intermediate. *Biochemistry*, 42:14501–14506, 2003.

[250] P. A. Malmqvist, A. Rendell and B. O. Roos. The restricted active space self-consistent-field method, implemented with a split graph unitary-group approach. *J. Phys. Chem.*, 94(14):5477–5482, 1990.

[251] J. Olsen, B. O. Roos, P. Jorgensen and H. J. A. Jensen. Determinant based configuration-interaction algorithms for complete and restricted configuration-interaction spaces. *J. Chem. Phys.*, 89(4):2185–2192, 1988.

[252] M. Boggio-Pasqua, M. A. Robb and M. J. Bearpark. Photostability via a sloped conical intersection: A casscf and rasscf study of pyracylene. *J. Phys. Chem. A*, 109(39):8849–8856, 2005.

[253] S. Olsen and S. C. Smith. Radiationless decay of red fluorescent protein chromophore models via twisted intermolecular charge transfer states. *J. Am. Chem. Soc.*, 129(7):2054–2065, 2007.

[254] J. J. van Thor, T. Gensch, K. J. Hellingwerf and L. N. Johnson. Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222. *Nat. Struct. Biol.*, 9(1):37–41, 2002.

[255] A. F. Bell, D. Stoner-Ma, R. M. Wachter and P. J. Tonge. Light-driven decarboxylation of wild-type green fluorescent protein. *J. Am. Chem. Soc.*, 125(23):6919–6926, 2003.

[256] D. Yarbrough, R. M. Wachter, K. Kallio, M. V. Matz and S. J. Remington. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0 Å resolution. *Proc. Natl. Acad. Sci. USA*, 98(2):462–467, 2001.

[257] R. Ando, H. Hama, M. Yamamoto-Hino, H. Mizuno and A. Miyawaki. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 99(20):12651–12656, 2002.

[258] H. Tsutsui, S. Karasawa, H. Shimizu, N. Nukina and A. Miyawaki. Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. *EMBO Rep.*, 6(3):233–238, 2005.

[259] J. Wiedenmann, A. Schenk, C. Röcker, A. Girod, K. D. Spindler and G. U. Nienhaus. A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from entacmaea quadricolor (anthozoa, actinaria). *Proc. Natl. Acad. Sci. USA*, 99(18):11646–11651, 2002.

[260] J. Petersen, P. G. Wilmann, T. Beddoe, A. J. Oakley, R. J. Devenish, M. Prescott and J. Rossjohn. The 2.0 Å crystal structure of eqFP611, a far red fluorescent protein from the sea anemone Entacmaea quadricolor. *J. Biol. Chem.*, 278(45):44626–44631, 2003.

[261] K. Nienhaus, B. Vallone, F. Renzi, J. Wiedenmann and G. U. Nienhaus. Crystallization and preliminary x-ray diffraction analysis of the red fluorescent protein eqFP611. *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, 59:1253–1255, 2003.

[262] D. C. Loos, S. Habuchi, C. Flors, J. I. Hotta, J. R. Wiedenmann, G. U. Nienhaus and J. Hofkens. Photoconversion in the red fluorescent protein from the sea anemone entacmaea quadricolor: Is cis-trans isomerization involved? *J. Am. Chem. Soc.*, 128(19):6270–6271, 2006.

[263] M. Prescott, M. Ling, T. Beddoe, A. J. Oakley, S. Dove, O. Hoegh-Guldberg, R. J. Devenish and J. Rossjohn. The 2.2 Å crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation. *Structure*, 11(3):275–284, 2003.

[264] P. G. Wilmann, J. Petersen, A. Pettikiriarachchi, A. M. Buckle, S. C. Smith, S. Olsen, M. A. Perugini, R. J. Devenish, M. Prescott and J. Rossjohn. The 2.1 Å crystal structure of the far-red fluorescent protein HcRed: Inherent conformational flexibility of the chromophore. *J. Mol. Biol.*, 349(1):223–237, 2005.

[265] H.-J. Werner et al. MOLPRO, Version 2002.6. Birmingham, UK, 2003.

[266] M. R. Redinbo, L. Stewart, P. Kuhn, J. J. Champoux and W. G. J. Hol. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science*, 279(5356):1504–1513, 1998.

[267] W. L. Jorgensen. Quantum and statistical mechanical studies of liquids. Transferable intermolecular potential functions for water, alcohols, and ethers - application to liquid water. *J. Am. Chem. Soc.*, 103(2):335–340, 1981.

[268] T. S. Rose, M. J. Rosker and A. H. Zewail. Femtosecond real-time observation of wave packet oscillations (resonance) in dissociation reactions. *J. Chem. Phys.*, 88(10):6672–6673, 1988.

[269] M. J. Rosker, T. S. Rose and A. H. Zewail. Femtosecond real-time dynamics of photofragment-trapping resonances on dissociative potential-energy surfaces. *Chem. Phys. Lett.*, 146(3-4):175–179, 1988.

[270] L. Blancafort, P. Hunt and M. A. Robb. Intramolecular electron transfer in bis(methylene) adamantyl radical cation: A case study of diabatic trapping. *J. Am. Chem. Soc.*, 127(10):3391–3399, 2005.

[271] A. Douhal, S. K. Kim and A. H. Zewail. Femtosecond molecular-dynamics of tautomerization in model base-pairs. *Nature*, 378(6554):260–263, 1995.

[272] P. R. L. Markwick, N. L. Doltsinis and J. Schlitter. Probing irradiation induced DNA damage mechanisms using excited state Car-Parrinello molecular dynamics. *J. Chem. Phys.*, 126(4):0451041–0451047, 2007.

[273] C. E. Crespo-Hernandez, B. Cohen and B. Kohler. Base stacking controls excited-state dynamics in A-T DNA. *Nature*, 436(7054):1141–1144, 2005.

[274] S. Shinkai and O. Manabe. Photocontrol of ion extraction and ion-transport by photofunctional crown ethers. *Topics in current Chemistry*, 121:67–104, 1984.

[275] R. H. Kramer, J. J. Chambers and D. Trauner. Photochemical tools for remote control of ion channels in excitable cells. *Nat. Chem. Biol.*, 1(7):360–365, 2005.

[276] H. Bayley. Ion channels get flashy. *Nat. Chem. Biol.*, 2(1):11–13, 2006.

[277] L. Ulysse, J. Cubillos and J. Chmielewski. Photoregulation of cyclic peptide conformation. *J. Am. Chem. Soc.*, 117(32):8466–8467, 1995.

[278] I. Willner. Photoswitchable biomaterials: En route to optobioelectronic systems. *Accounts Chem. Res.*, 30(9):347–356, 1997.

[279] H. Asanuma, T. Ito, T. Yoshida, X. G. Liang and M. Komiyama. Photoregulation of the formation and dissociation of a DNA duplex by using the cis-trans isomerization of azobenzene. *Angew. Chem.-Int. Edit.*, 38(16):2393–2395, 1999.

[280] S. Spörlein, H. Carstens, H. Satzger, C. Renner, R. Behrendt, L. Moroder, P. Tavan, W. Zinth and J. Wachtveitl. Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA*, 99(12):7998–8002, 2002.

[281] V. Kräutler, A. Aemissegger, P. H. Hünenberger, D. Hilvert, T. Hansson and W. F. van Gunsteren. Use of molecular dynamics in the design and structure determination of a photoinducible beta-hairpin. *J. Am. Chem. Soc.*, 127(13):4935–4942, 2005.

[282] A. Aemissegger, V. Kräutler, W. F. van Gunsteren and D. Hilvert. A photoinducible beta-hairpin. *J. Am. Chem. Soc.*, 127(9):2929–2936, 2005.

[283] L. Guerrero, O. S. Smart, C. J. Weston, D. C. Burns, G. A. Woolley and R. K. Allemann. Photochemical regulation of DNA-binding specificity of MyoD. *Angew. Chem.-Int. Edit.*, 44:7778–7782, 2005.

[284] P. H. Nguyen, Y. G. Mu and G. Stock. Structure and energy landscape of a photoswitchable peptide: A replica exchange molecular dynamics study. *Proteins*, 60(3):485–494, 2005.

[285] P. H. Nguyen and G. Stock. Nonequilibrium molecular dynamics simulation of a photoswitchable peptide. *Chem. Phys.*, 323(1):36–44, 2006.

[286] S. Yasuda, T. Nakamura, M. Matsumoto and H. Shigekawa. Phase switching of a single isomeric molecule and associated characteristic rectification. *J. Am. Chem. Soc.*, 125(52):16430–16433, 2003.

[287] J. Henzl, M. Mehlhorn, H. Gawronski, K. H. Rieder and K. Morgenstern. Reversible cis-trans isomerization of a single azobenzene molecule. *Angew. Chem.-Int. Edit.*, 45(4):603–606, 2006.

[288] S. Zou, M. A. Hempenius, H. Schönherr and G. J. Vancso. Force spectroscopy of individual stimulus-responsive poly(ferrocenyldimethylsilane) chains: towards a redox-driven macromolecular motor. *Macromolecular Rapid Communications*, 27:103–108, 2006.

[289] T. Schultz, J. Quenneville, B. Levine, A. Toniolo, T. J. Martinez, S. Lochbrunner, M. Schmitt, J. P. Shaffer, M. Z. Zgierski and A. Stolow. Mechanism and dynamics of azobenzene photoisomerization. *J. Am. Chem. Soc.*, 125(27):8098–8099, 2003.

[290] C. Nonnenberg, H. E. Gaub and I. Frank. First-principles simulation of the photoreaction of a capped azobenzene: the rotational pathway is feasible. *ChemPhysChem*, 7:1455–1461, 2006.

[291] S. Monti, G. Orlandi and P. Palmieri. Features of the photochemically active state surfaces of azobenzene. *Chem. Phys.*, 71(1):87–99, 1982.

[292] P. Cattaneo and M. Persico. An ab initio study of the photochemistry of azobenzene. *Phys. Chem. Chem. Phys.*, 1(20):4739–4743, 1999.

[293] C. Ciminelli, G. Granucci and M. Persico. The photoisomerization mechanism of azobenzene: A semiclassical simulation of nonadiabatic dynamics. *Chem. Eur. J.*, 10:2327–2341, 2004.

[294] A. Cembran, F. Bernardi, M. Garavelli, L. Gagliardi and G. Orlandi. On the mechanism of the cis-trans isomerization in the lowest electronic states of azobenzene: S-0, S-1, and T-1. *J. Am. Chem. Soc.*, 126(10):3234–3243, 2004.

[295] C. R. Crecca and A. E. Roitberg. Theoretical study of the isomerization mechanism of azobenzene and disubstituted azobenzene derivatives. *J. Phys. Chem. A*, 110(26):8188–8203, 2006.

[296] C. T. Lee, W. T. Yang and R. G. Parr. Development of the collesalvetcorrelation-energy formula into a functional of the electron-density. *Phys. Rev. B*, 37(2):785–789, 1988.

[297] A. A. Blevins and G. J. Blanchard. Effect of positional substitution on the optical response of symmetrically disubstituted azobenzene derivatives. *J. Phys. Chem. B*, 108(16):4962–4968, 2004.

[298] W. L. DeLano. *The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org*, 2002.

[299] C. Bustamante, J. F. Marko, E. D. Siggia and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994.

[300] R. Behrendt, C. Renner, M. Schenk, F. Q. Wang, J. Wachtveitl, D. Oesterhelt and L. Moroder. Photomodulation of the conformation of cyclic peptides with azobenzene moieties in the peptide backbone. *Angew. Chem.-Int. Edit.*, 38(18):2771–2774, 1999.

[301] G. Neuert, T. Hugel, R. R. Netz and H. E. Gaub. Elasticity of poly(azobenzene-peptides). *Macromolecules*, 39(2):789–797, 2006.

[302] S. Habuchi, P. Dedecker, J. I. Hotta, C. Flors, R. Ando, H. Mizuno, A. Miyawaki and J. Hofkens. Photo-induced protonation/deprotonation in the GFP-like fluorescent protein Dronpa: mechanism responsible for the reversible photoswitching. *Photochem. Photobiol. Sci.*, 5(6):567–576, 2006.

[303] E. Fron, C. Flors, G. Schweitzer, S. Habuchi, R. Ando, H. Mizuno, F. C. De Schryver, A. Miyawaki and J. Hofkens. Ultrafast excited-state dynamics of the photoswitchable protein Dronpa. *J. Am. Chem. Soc.*, 129(16):4870–4871, 2007.

Lebenslauf von Lars Schäfer

Persönliche Daten

Name: Lars Schäfer
Adresse: Auf der Lieth 18, 37077 Göttingen
Geburtsdatum, -ort: 16.03.1978, Braunschweig
Nationalität: Deutsch

Ausbildung

06.1997	Abitur an der Neuen Oberschule, Braunschweig Prüfungsfächer: Chemie, Biologie, Englisch, Erdkunde
09.1997 – 09.1998	Zivildienst
10.1998 – 09.2003	Technische Universität Carolo-Wilhelmina zu Braunschweig Immatrikuliert für Diplomstudiengang Chemie
10.2000	Vordiplom in Chemie Gesamtnote: sehr gut
11.2003	Diplom in Chemie Gesamtnote: Mit Auszeichnung
10.2003 – 06.2007	Max-Planck-Institut für Biophysikalische Chemie, Göttingen. Doktorarbeit mit dem Titel: Photoactivated Processes in Condensed Phase studied by Molecular Dynamics Simulations. Betreuer: Prof. Helmut Grubmüller Note: Mit Auszeichnung